

CRFE

SEARCH REQUEST FORM

meq

Scientific and Technical Information Center

Requester's Full Name: John Hines Examiner #: _____ Date: 5/3/01
Art Unit: 1045 Phone Number: 305-0481 Serial Number: 07926776
Mail Box and Bldg/Room Location: 7E12 Results Format Preferred (circle) PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: isolated nucleic acids

Inventors (please provide full names): JEAN LING DING DEJIAN SONG TAN

Bow HO

Earliest Priority Filing Date: 10/24/99

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search SEQ ID NO 10 +
SEQ ID NO 11

Manuscript

meq 11
pmt 10

STAFF USE ONLY

Type of Search

Vendors and cost where applicable

Searcher: Patricia J. Jolly NA Sequence (#) 1 STN _____
Searcher Phone #: 305-0481 AA Sequence (#) _____ Dialog _____
Searcher Location: Patricia J. Jolly Structure (#) _____ Questia/Orbit _____
Date Searcher Picked Up: 5/3/01 Bibliographic _____ Dr. Link _____
Date Completed: 5/19/01 Litigation _____ Lexis/Nexis _____
Searcher Prep & Review Time: _____ Fulltext _____ Sequence Systems: Access _____
Clerical Prep Time: 3 min Patent Family _____ WWW/Internet _____
Online Time: 2 min Other _____ Other (specify) _____

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2001, 15:51:41 : Search time 17.84 Seconds
(without alignments)
137.969 Million cell updates/sec

Title: US-09-426-776-10

Perfect score: 96

Sequence: 1 MRVLVLAVALVGGGSLG 21

Scoring table: BLOSUM62 Gapop 10.0, Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL15:

1: sp_archaea:

2: sp_bacteria:

3: sp_fungi:

4: sp_human:

5: sp_invertebrate:

6: sp_mammal:

7: sp_mmc:

8: sp_organelle:

9: sp_plant:

10: sp_protist:

11: sp_recent:

12: sp_unclassified:

13: sp_vertebrate:

14: sp_virus:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	85.4	1788	13	09YCK0
2	48	50.0	330	10	09XIV8
3	45	46.9	176	5	09W4U2
4	45	46.9	262	10	09LV78
5	45	46.9	330	5	018118
6	45	46.9	691	3	09UT43
7	45	46.9	1718	5	09VGB4
8	44	45.8	328	2	09RR76
9	44	45.8	723	1	P96433
10	44	45.8	815	5	044386
11	44	45.8	870	2	09W445
12	44	45.8	992	2	051918
13	44	45.8	1108	5	09U280
14	44	45.8	1297	5	09V6P5
15	44	45.3	952	5	097198
16	43.5	45.3	952	5	097198
17	43	44.8	127	2	09RTH6
18	43	44.8	181	4	09NTL2
19	43	44.8	273	2	006170

20	43	44.8	346	2	09PCH7
21	43	44.8	359	5	09Y179
22	43	44.8	359	1	047973
23	43	44.8	583	2	09RY54
24	43	44.8	598	2	083032
25	43	44.8	599	2	09L010
26	43	44.8	613	4	09NTM3
27	43	44.8	729	4	09NT12
28	43	44.8	767	5	09V926
29	42.5	44.3	893	5	016221
30	42.5	44.3	1113	2	09L249
31	42	43.8	185	5	09NXL8
32	42	43.8	107	10	09LE87
33	42	43.8	191	1	09Y455
34	42	43.8	191	1	09S531
35	42	43.8	297	1	029031
36	42	43.8	338	2	09X595
37	42	43.8	460	2	006312
38	42	43.8	634	2	09S658
39	42	43.8	657	2	086463
40	42	43.8	1339	13	09W612
41	41.5	43.2	296	5	09V729
42	41.5	43.2	351	2	09PL26
43	41.5	43.2	463	3	09P422
44	41.5	43.2	1363	3	09P422
45	41	42.7	109	2	09RNX6

ALIGNMENTS

RESULT	1
09YCK0	PRELIMINARY; PRT: 1788 AA.
ID	09YCK0
AC	09YCK0
DT	01-MAY-1999 (TREMBLrel. 10, Created)
DR	01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE	01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE	VITELLOGENIN PRECURSOR.
GN	VTGL
OS	Oreochromis aureus.
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC	Acanthomorpha; Acanthopterygii; Perciformes; Labroidel;
OC	Cichlidae; Oreochromis.
OX	NCBI_TaxID=47969;
RN	SEQUENCE FROM N.A.
RP	111
RC	TISSUE=Liver;
RT	Cloning of Full-length Oreochromis aureus Vitellogenin cDNA and its
RT	Deduced Primary Structure.
RL	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

RN	SEQUENCE FROM N.A.
RP	111
RC	TISSUE=Blood;
RT	Cloning of Full-length Oreochromis aureus Vitellogenin cDNA and its
RT	Deduced Primary Structure.
RL	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

RN	SEQUENCE FROM N.A.
RP	111
RC	TISSUE=Blood;
RT	Cloning of Full-length Oreochromis aureus Vitellogenin cDNA and its
RT	Deduced Primary Structure.
RL	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

RN	SEQUENCE FROM N.A.
RP	111
RC	TISSUE=Blood;
RT	Cloning of Full-length Oreochromis aureus Vitellogenin cDNA and its
RT	Deduced Primary Structure.
RL	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

RN	SEQUENCE FROM N.A.
RP	111
RC	TISSUE=Blood;
RT	Cloning of Full-length Oreochromis aureus Vitellogenin cDNA and its
RT	Deduced Primary Structure.
RL	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

Conservation

Query Match 85.4% Score 82; DB 13; Length 1788;
 Best Local Similarity 95.0%; Pred. No. 0.0031; Mismatches 1; Indels 0; Gaps 0;
 Matches 19; Conservative 0;
 QY 1 MRVYLALAVLAVGDSNLG 20
 DB 1 MRVYLALAVLAVGDSNLG 20
 RESULT 2
 Q9XIV8 PRELIMINARY; PRT; 330 AA.
 AC Q9XIV8;
 DT 01-NOV-1999 (TRENBLrel. 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
 DE PEROXIDASE (EC 1.11.1.7).
 OS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
 OC Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. SAMSON NN; TISSUE=TMV-INFECTED LEAF;
 RA Hiraga S., Ito H., Matsui H., Honma M., Ohashi Y.;
 RT CDNA Sequences for Two Novel Tobacco Peroxidase Isoenzymes.*;
 RV EMBL: AB021941 (1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB021941; GenBank: AF42307.1;
 DR HSP: P00433; JATU;
 DR INTERPRO: IPR000823;
 DR INTERPRO: IPR002016;
 DR INTERPRO: IPR002207;
 DR PFAM: PF00141; peroxidase; 1.
 DR PRINTS: PR00458; PEROXIDASE.
 DR PRINTS: PR00459; ASPEROXIDASE.
 DR PRINTS: PR00461; PEROXIDASE.
 DR PROSITE: PS00435; PEROXIDASE_1; 1.
 DR PROSITE: PS00435; PEROXIDASE_2; 1.
 DR PROSITE: PS00435; PEROXIDASE.
 KW Peroxidase; Oxidoreductase.
 SQ SEQUENCE 330 AA; 35731 MW; 2034AC56376B6180 CRC64;

Query Match 50.0% Score 48; DB 10; Length 330;
 Best Local Similarity 42.9%; Pred. No. 34;
 Matches 9; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
 QY 1 MRVYLALAVLAVGDSNLG 21
 DB 15 MRVYLALAVLAVGDSNLG 35

RESULT 3
 Q9W4U2 PRELIMINARY; PRT; 176 AA.
 AC Q9W4U2;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
 DE G3603 PROTEIN.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY.
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Belmont L., Bohnert A., Bohnert S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Branton R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA An H.-J., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Ballou R.M., Bayraktaroglu L., Bayraktaroglu L., Baldwin D.,
 RA Beeson K.Y., Benson P.V., Bock J., Bock J., Bock J., Bock J.,
 RA Borkova D., Borkova D., Borkova D., Borkova D., Borkova D.,
 RA Burtis J.C., Busan D.A., Butler H., Cadieu F., Cantor M.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
 RA Hostler S., Howe K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Jaiswal M., Kalush E., Kaul R., Kaul R., Kaul R., Kaul R.,
 RA Kimmel B.E., Kodira C.D., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Li Z., Li Z.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Schmeckel K., Staden-Klamis I., Simpson M., Skupski M.P., Smith T.,
 RA Spier F., St. Tsching C., Stappleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yan S., Yao L.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhou L.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT The genome sequence of Drosophila melanogaster.*;
 DL Science 287:2185-2195 (2000).
 DR HSP: P007426; AF45844.1;
 DR FLXBASE: FB00029648; CG3603.
 DR INTERPRO: IPR000276;
 DR INTERPRO: IPR002198;
 DR INTERPRO: IPR002347;
 DR PFAM: PF00106; adh_short; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 DR PROSITE: PS00237; G_PROTEIN_RECEPTOR; UNKNOWN_1.
 SQ SEQUENCE 176 AA; 18567 MW; 4F0D8EAAEA294324 CRC64;

Query Match 45.9% Score 45; DB 5; Length 176;
 Best Local Similarity 57.1%; Pred. No. 48;
 Matches 12; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
 QY 1 MRVYLALAVLAVGDSNLG 21
 DB 1 MSVGLKALVAVTGSGIG 21

RESULT 4
 Q9LV78 PRELIMINARY; PRT; 262 AA.
 AC Q9LV78;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
 DE GENOMIC DNA, CHROMOSOME 5, PL CLONE:MKK3.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Rosales; Rosaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA:
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RT Tabata S.
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty pl and TAC
RT clones." J. 31-63 (2000)
RT DNA Res. 9:123-6 (1999)
RT DIAL. 10:123-6 (1999)
RT SOURCE: 262 AB. 5CD937647893RALL C9667
SO NCBI

Query Match	46.9%	Score 45;	DB 10;	Length 262;
Best Local Similarity	56.2%	Pred. No. 72;		
Matches 9;	Conservative 4;	Mismatches 3;	Indels 0;	Gaps 0;

Qy 4 LVLALAVALAVDGSN 19
|::|||::|||
Db 82 LLMAAVLLSVGEGSN 97

RESULT 5
O18118
ID O18118
PRELIMINARY;

AC	DT	DT	DT	DT
	01-JAN-1998	(TRENDEL. 05, Created)		
	01-JAN-1998	(TRENDEL. 05, Last sequence update)		
	01-JAN-1999	(TRENDEL. 09, Last annotation update)		

DE 123Fl.0 PROTEIN.
GN T23Fl.6.
OS *Caenorhabditis elegans*.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; R

OC Knauliaceae; Perideridae; Caenofabriditis.
OX NCBI_TaxID=6239;
RN {1}
RP SEQUENCE FROM N.A.

RA Wilkinson J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RD SEQUENCE FROM N A

RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Garter M., Dean S., Dushkin P., Fawcett J., Fawcett J., Fawcett J.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA

RA Snaaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;

RT *elegans*.^a;
RL Nature 368:32-38(1994).
DR EMBL; Z81129; CAB03405.1; -

Query Match **46.9%** **Score 45** **DB 5** **Length 330**;

Qv 1 MRVLVLALAVAVGDSNL 20
Matches 9; Conservative 5; Mismatches. 6; Indels 0; Gaps

Db 1 MRCVILFAVALAVAQASSI 20

RESULT 6
Q90T43
ID Q90T43
PRELIMINARY; PRT; 691 AA.
AC Q90T43.

DT	01-MAY-2000	(TREMBLrel. 13, Created)
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)
DT	01-JUN-2000	(TREMBLrel. 14, Last annotation update)

DE	POTUTIVE ATPASE (FRAGMENT).	
GN	SPRC821.113C.	
OS	Schizosaccharomyces pombe (Fission yeast).	
OC	Ascomycota Fungi.	
NC	Schizosaccharomycetales;	
NC	Schizosaccharomycetaceae; Schizosaccharomycetes.	
OX	NB1_TaxID=4896;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=972H;	
RA	Rieger M., Lyne M., Rajandream A.A., Barrell B.G.;	
RL	Submitted (OCT-1995) to the EMBL/GenBank/DBSJ databases.	
RL	EMBL: AL11770; CAS57447.1;	
DR	INTERPRO: IPR001757;	
DR	PFAM: PF00112; E1-B2.ATPase; 1.	
FT	NON_TER 691	
FT	SEQUENCE 691 AA; 78275 MW; F0B8FEAA5753CC23 CRC64;	

Query Match 46.9%; Score 45; DB 3; Length 691;
Best Local Similarity 64.3%; Pred. No. 1.9e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy	6 LALAYALAVGDSN 19
Db	362 LEKAVTLAIGDGN 375

RESULT	7
Q9VGD4	

ID	Q3VGD4	FREELIMINARI;	FRI;	1716 MO.
AC	Q9VGD4;			
DT	01-MAY-2000	(TREMBLrel.	13,	Created)
DT	01-MAY-2000	(TREMBLrel.	13,	Last sequence update)

DI 01-JUN-2000 (FEBLEFRIE: 14, Last annotation update)
DE CG14741 PROTEIN.
GN CG14741.
OS *Drosophila melanogaster* (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI TAXID:7227.

RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=BERKELEY;	
SV	WPPY TUB-20105005; PubMed-10731122.	

RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.

RA Jadaill M., Natusi F., Kalpen G.N., Ae Z., Kemnison J.A., Reclum A.G.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Lin X., Martel R., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarly C., Morris J., Moscarel A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazuelos M., Pittman C.S., Dan G., Pollard T., Burl V., Boese M. G.

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

Spur B.C., Sidani-Klaonis I., Simpson M., Skupski M.P., Smith T.,
Spur E., Spradling C., Starr-Gordon E., Su B., Sun M.,
Svavarskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.-S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong P.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.N., Rubin G.M., Venter J.C.,
*The genome sequence of *Drosophila melanogaster*.*
Science 287:2185-2195(2000).
EMBL: AF003694; AAF54749.1; -
FLYBASE: FBgn0037989; CG14741.
INTERPRO: IPR000661; -
Pfam: PF00122; PF00123; PF00124; Pfam0152; Pfam0153; Pfam0154; Pfam0155; Pfam0156; Pfam0157; Pfam0158; Pfam0159; Pfam0160; Pfam0161; Pfam0162; Pfam0163; Pfam0164; Pfam0165; Pfam0166; Pfam0167; Pfam0168; Pfam0169; Pfam0170; Pfam0171; Pfam0172; Pfam0173; Pfam0174; Pfam0175; Pfam0176; Pfam0177; Pfam0178; Pfam0179; Pfam0180; Pfam0181; Pfam0182; Pfam0183; Pfam0184; Pfam0185; Pfam0186; Pfam0187; Pfam0188; Pfam0189; Pfam0190; Pfam0191; Pfam0192; Pfam0193; Pfam0194; Pfam0195; Pfam0196; Pfam0197; Pfam0198; Pfam0199; Pfam0200; Pfam0201; Pfam0202; Pfam0203; Pfam0204; Pfam0205; Pfam0206; Pfam0207; Pfam0208; Pfam0209; Pfam0210; Pfam0211; Pfam0212; Pfam0213; Pfam0214; Pfam0215; Pfam0216; Pfam0217; Pfam0218; Pfam0219; Pfam0220; Pfam0221; Pfam0222; Pfam0223; Pfam0224; Pfam0225; Pfam0226; Pfam0227; Pfam0228; Pfam0229; Pfam0230; Pfam0231; Pfam0232; Pfam0233; Pfam0234; Pfam0235; Pfam0236; Pfam0237; Pfam0238; Pfam0239; Pfam0240; Pfam0241; Pfam0242; Pfam0243; Pfam0244; Pfam0245; Pfam0246; Pfam0247; Pfam0248; Pfam0249; Pfam0250; Pfam0251; Pfam0252; Pfam0253; Pfam0254; Pfam0255; Pfam0256; Pfam0257; Pfam0258; Pfam0259; Pfam0260; Pfam0261; Pfam0262; Pfam0263; Pfam0264; Pfam0265; Pfam0266; Pfam0267; Pfam0268; Pfam0269; Pfam0270; Pfam0271; Pfam0272; Pfam0273; Pfam0274; Pfam0275; Pfam0276; Pfam0277; Pfam0278; Pfam0279; Pfam0280; Pfam0281; Pfam0282; Pfam0283; Pfam0284; Pfam0285; Pfam0286; Pfam0287; Pfam0288; Pfam0289; Pfam0290; Pfam0291; Pfam0292; Pfam0293; Pfam0294; Pfam0295; Pfam0296; Pfam0297; Pfam0298; Pfam0299; Pfam0300; Pfam0301; Pfam0302; Pfam0303; Pfam0304; Pfam0305; Pfam0306; Pfam0307; Pfam0308; Pfam0309; Pfam0310; Pfam0311; Pfam0312; Pfam0313; Pfam0314; Pfam0315; Pfam0316; Pfam0317; Pfam0318; Pfam0319; Pfam0320; Pfam0321; Pfam0322; Pfam0323; Pfam0324; Pfam0325; Pfam0326; Pfam0327; Pfam0328; Pfam0329; Pfam0330; Pfam0331; Pfam0332; Pfam0333; Pfam0334; Pfam0335; Pfam0336; Pfam0337; Pfam0338; Pfam0339; Pfam0340; Pfam0341; Pfam0342; Pfam0343; Pfam0344; Pfam0345; Pfam0346; Pfam0347; Pfam0348; Pfam0349; Pfam0350; Pfam0351; Pfam0352; Pfam0353; Pfam0354; Pfam0355; Pfam0356; Pfam0357; Pfam0358; Pfam0359; Pfam0360; Pfam0361; Pfam0362; Pfam0363; Pfam0364; Pfam0365; Pfam0366; Pfam0367; Pfam0368; Pfam0369; Pfam0370; Pfam0371; Pfam0372; Pfam0373; Pfam0374; Pfam0375; Pfam0376; Pfam0377; Pfam0378; Pfam0379; Pfam0380; Pfam0381; Pfam0382; Pfam0383; Pfam0384; Pfam0385; Pfam0386; Pfam0387; Pfam0388; Pfam0389; Pfam0390; Pfam0391; Pfam0392; Pfam0393; Pfam0394; Pfam0395; Pfam0396; Pfam0397; Pfam0398; Pfam0399; Pfam0400; Pfam0401; Pfam0402; Pfam0403; Pfam0404; Pfam0405; Pfam0406; Pfam0407; Pfam0408; Pfam0409; Pfam0410; Pfam0411; Pfam0412; Pfam0413; Pfam0414; Pfam0415; Pfam0416; Pfam0417; Pfam0418; Pfam0419; Pfam0420; Pfam0421; Pfam0422; Pfam0423; Pfam0424; Pfam0425; Pfam0426; Pfam0427; Pfam0428; Pfam0429; Pfam0430; Pfam0431; Pfam0432; Pfam0433; Pfam0434; Pfam0435; Pfam0436; Pfam0437; Pfam0438; Pfam0439; Pfam0440; Pfam0441; Pfam0442; Pfam0443; Pfam0444; Pfam0445; Pfam0446; Pfam0447; Pfam0448; Pfam0449; Pfam0450; Pfam0451; Pfam0452; Pfam0453; Pfam0454; Pfam0455; Pfam0456; Pfam0457; Pfam0458; Pfam0459; Pfam0460; Pfam0461; Pfam0462; Pfam0463; Pfam0464; Pfam0465; Pfam0466; Pfam0467; Pfam0468; Pfam0469; Pfam0470; Pfam0471; Pfam0472; Pfam0473; Pfam0474; Pfam0475; Pfam0476; Pfam0477; Pfam0478; Pfam0479; Pfam0480; Pfam0481; Pfam0482; Pfam0483; Pfam0484; Pfam0485; Pfam0486; Pfam0487; Pfam0488; Pfam0489; Pfam0490; Pfam0491; Pfam0492; Pfam0493; Pfam0494; Pfam0495; Pfam0496; Pfam0497; Pfam0498; Pfam0499; Pfam0500; Pfam0501; Pfam0502; Pfam0503; Pfam0504; Pfam0505; Pfam0506; Pfam0507; Pfam0508; Pfam0509; Pfam0510; Pfam0511; Pfam0512; Pfam0513; Pfam0514; Pfam0515; Pfam0516; Pfam0517; Pfam0518; Pfam0519; Pfam0520; Pfam0521; Pfam0522; Pfam0523; Pfam0524; Pfam0525; Pfam0526; Pfam0527; Pfam0528; Pfam0529; Pfam0530; Pfam0531; Pfam0532; Pfam0533; Pfam0534; Pfam0535; Pfam0536; Pfam0537; Pfam0538; Pfam0539; Pfam0540; Pfam0541; Pfam0542; Pfam0543; Pfam0544; Pfam0545; Pfam0546; Pfam0547; Pfam0548; Pfam0549; Pfam0550; Pfam0551; Pfam0552; Pfam0553; Pfam0554; Pfam0555; Pfam0556; Pfam0557; Pfam0558; Pfam0559; Pfam0560; Pfam0561; Pfam0562; Pfam0563; Pfam0564; Pfam0565; Pfam0566; Pfam0567; Pfam0568; Pfam0569; Pfam0570; Pfam0571; Pfam0572; Pfam0573; Pfam0574; Pfam0575; Pfam0576; Pfam0577; Pfam0578; Pfam0579; Pfam0580; Pfam0581; Pfam0582; Pfam0583; Pfam0584; Pfam0585; Pfam0586; Pfam0587; Pfam0588; Pfam0589; Pfam0590; Pfam0591; Pfam0592; Pfam0593; Pfam0594; Pfam0595; Pfam0596; Pfam0597; Pfam0598; Pfam0599; Pfam0600; Pfam0601; Pfam0602; Pfam0603; Pfam0604; Pfam0605; Pfam0606; Pfam0607; Pfam0608; Pfam0609; Pfam0610; Pfam0611; Pfam0612; Pfam0613; Pfam0614; Pfam0615; Pfam0616; Pfam0617; Pfam0618; Pfam0619; Pfam0620; Pfam0621; Pfam0622; Pfam0623; Pfam0624; Pfam0625; Pfam0626; Pfam0627; Pfam0628; Pfam0629; Pfam0630; Pfam0631; Pfam0632; Pfam0633; Pfam0634; Pfam0635; Pfam0636; Pfam0637; Pfam0638; Pfam0639; Pfam0640; Pfam0641; Pfam0642; Pfam0643; Pfam0644; Pfam0645; Pfam0646; Pfam0647; Pfam0648; Pfam0649; Pfam0650; Pfam0651; Pfam0652; Pfam0653; Pfam0654; Pfam0655; Pfam0656; Pfam0657; Pfam0658; Pfam0659; Pfam0660; Pfam0661; Pfam0662; Pfam0663; Pfam0664; Pfam0665; Pfam0666; Pfam0667; Pfam0668; Pfam0669; Pfam0670; Pfam0671; Pfam0672; Pfam0673; Pfam0674; Pfam0675; Pfam0676; Pfam0677; Pfam0678; Pfam0679; Pfam0680; Pfam0681; Pfam0682; Pfam0683; Pfam0684; Pfam0685; Pfam0686; Pfam068

```

Query Match      46.94; Score 45; DB 5; Length 1718;
Best Local Similarity 45.84; Pred.No. 4.7e+07;
Matches 11; Conservative 4; Mismatches 3; Indels 6; Gaps 1;

Qy      2 RVLTVAL-----AVAAVGDSGN 19
          : 11:1
          ||:||||:
DB      1270 KALVVELIKRANAVTALIGGAN 1293

```

RESULT	0
Q98R76	
ID	PRELIMINARY; PRT; 328 AA.
QC	Q98R76:
DT	01-MAY-2000 (TREMBlrel. 13, Created)
DD	01-MAY-2000 (TREMBlrel. 13, Last sequence update)
OT	01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE	GLYCEROL-3-PHOSPHATE DEHYDROGENASE, NAD(+)-DEPENDENT.
GN	Df621.
OGS	Deinococcus radiodurans.
OC	Bacteria: Thermus/Deinococcus group: Deinococcales: Deinococcus.
CC	NCBI_TaxID=1299;
RRN	(1)
RNR	SEQUENCE FROM NA.
RNC	STRAIN=RI-20036896; PubMed=10557265;
RNX	White, R., Elsdon, J., Haffner, G., Hickey, E.K., Peterson, J.D.,
RAA	Moffat K.S., Qip H. Jiang, W. Nelson, W.C. Albracht D.L.,
RAB	Doddart K.S., Qip H. Jiang, L. Pamphile W. Crosby M. Shen M.,
RAC	Matkova K.J., Lam P., McDonald T., O'tterback T., Zaleski C.,
RAD	Amakharava S.S., Aravind L., Daly M.J., Antonick K.W., Fleischmann R.D.,
RAE	Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RAF	Fraser C.M.;
RAG	*Genome Sequence of the Radioresistant Bacterium Deinococcus
RAT	radiodurans RI.*;
RFT	Science 286:1571-1577(1999).
RRL	EMBL: AE002091; AAF12158.1; -
RRC	TIGR: Df621.; -
RRE	INTERPRO: IPR001652; -
DRD	INTERPRO: IPR001993; -
DRR	INTERPRO: ITC000000; -
DRS	PFAM PF01210; NAD_Glyc_dh_1.
DRU	Pfam: PF000077; GPDHCRNASE
DRV	PRINTS: PS00077; MITOCH_CARRIER UNKNOWN 1.
DRW	PROSITE: PS00215; NITOX_LIGASE IL 2: UNKNOWN 1.
DRX	PROSITE: PS00339; AA.TRNA_LIGASE IL 2: UNKNOWN 1.
DRY	SEQUENCE 328 AA; 33597 MW: 46D909D7D14170BE CRC64:
DRZ	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	
DR2	
DR3	
DR4	
DR5	
DR6	
DR7	
DR8	
DR9	
DR0	
DR1	

```

Query Match      45.8%  Score 44;  DB 2;  Length 328;
Best Local Similarity 52.6%  Pred. NO. 1.2e+02;
Matches 10;  Conservative 3;  Mismatches 6;  Indels 0;  Gaps 0;

```

DBD		184	VKLVNINACROGDLHLS	202
RESULT	9			
P69493				
AD	P69493		PRELIMINARY;	
ID	P69493:			
IC	01-MAY-1997	(T)EMBLrel.	03.	C
DT	01-MAY-1997	(T)EMBLrel.	03.	I
OT	01-MAY-2000	(T)EMBLrel.	13.	I
DE	PUTATIVE GLYCINE DEHYDROGENASE (EC 1.4.4.2) (GLYCINE DECARBOXYLASE CYCLIC CLEAVAGE SYSTEM P-PHOSPHATE) QNRJ.			
GN	Thermus aquaticus (subs. thermophilus)			
NC	U00096; Thermus/Delinoecoccus			
XX	NB1_0410-274;			
OR	SEQUENCE FROM N.A.			
RCP	STRAIN-H27;			
RA	Baetens M., Van de Casteele B, Lambermont Submitted (Feb-1997) to the EMBL			
CC	-!- CATALYTIC ACTIVITY: GLYCINE AMINOMETHYLDIHYDROLIPOLYMER + H ₂ O = COFATOR + AMINOETHYL-DIHYDROLIPOLYMER			
CC	EMBL: Y11457; CAAT2254.1;			
CD	PROSITE: PS01036; HSPT70.3; UN-			
KW	Oxidoreductase.			
FT	NON TER			
PP	SEQUENCE 333 AA; 35503 MW;			

```

Query Match          45.8%;
Best Local Similarity 60.0%;
Matches             9; Conservative

```

RESULT	10
PROG	PRELIMINARY;
ID	ID_090626;
ACC	ACC_090626;
DATE	01-MAY-2000 (T)EMBLrel. 13.
DT	01-MAY-2000 (T)EMBLrel. 13.
DD	01-OCT-2000 (T)EMBLrel. 15.
DE	CARTILAGE OLIGOMERIC MATRIX
DD	COMP.
GN	Mus musculus (Mouse).
GC	Eukaryota; Metazoa; Chordata
OC	Mammalia; Eutheria; Rodentia
OM	NCBI_TaxId=10090;
RC	SEQUENCE FROM N.A.
RT	TISSUE=CARTILAGE;
RA	Fang C., Carlson C.S., Loslie
RA	N.I.L., Di Cesare P.E.;
RT	Molecular Cloning, Sequencing,
RT	Mouse Cartilage Oligomeric M
RT	J. Orthop. Res. 0:0-0(1999).
DR	EMBL; AF013530; AAD01972.1;
DR	HSSP; F35555; ILMN.
DR	MGF:BB8469; Comp.
DR	INTERPRO: IPRO00561; -
DR	INTERPRO: IPRO01861; -
DR	PROSITE: PS00018; EXP 2; U
DR	PROSITE: PS00018; EXP HAND; U
DR	PROSITE: PS01187; EXP CA. 2.
DR	PROSITE: PS01187; EXP CA. 2.

KW Signal: Matrix protein.
 FT SIGNAL 1 19
 FT CHAIN 20 755
 SQ SEQUENCE 755 AA: 82352 MW: 70DFCF443589A0B7 CRC64;
 Query Match 45.88; Score 44; DB 11; Length 755;
 Best Local Similarity 66.74; Pred. No. 2.8e+02;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 3 VLVLALVALAVGQG 17
 DB 7 VLVLALVALAVGQG 21
 RESULT 11
 ID Q44385 PRELIMINARY; PRT: 815 AA.
 AC Q44385
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DE PUTATIVE E1-E2 ATPASE (FRAGMENT).
 DE POTATIVE E1-E2 ATPASE (FRAGMENT).
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 14028;
 RA Townsends S.M., Ficht T.A., Adams L.G., Baumber A.J.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DE Distribution of bcf operon in Salmonella enterica.*;
 DE EMBL; AF130422; AAD34372.1;
 DE EMBL; AF129435; AAD22098.1;
 DE INTERPRO: IPR000015;
 DE PFAM: PF00577; Usher; 1.
 DR PROSITE: PS01151; FIMBRIAL_USHER; 1.
 SQ SEQUENCE 815 AA: 91486 MW: D5B54ED4BDFDF232 CRC64;
 Query Match 45.88; Score 44; DB 5; Length 815;
 Best Local Similarity 64.34; Pred. No. 3.1e+02;
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 6 LALVALAVGDSN 19
 DB 431 LAKHVYLAIGDGN 444
 RESULT 12
 ID Q9W445 PRELIMINARY; PRT: 870 AA.
 AC Q9W445
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DE FIMBRIAL USHER.
 GN BCFC.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 14028;
 RA Townsends S.M., Ficht T.A., Adams L.G., Baumber A.J.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DE Distribution in mice does not predict the ability of Salmonella
 typhimurium to cause enteric fever.
 DE EMBL; AF130422; AAD34372.1;
 DE EMBL; AF129435; AAD22098.1;
 DE INTERPRO: IPR000015;
 DE PFAM: PF00577; Usher; 1.
 DR PROSITE: PS01151; FIMBRIAL_USHER; 1.
 SQ SEQUENCE 815 AA: 91486 MW: D5B54ED4BDFDF232 CRC64;

RC STRAIN-ATCC 14028;
 RA Townsends S.M., Ficht T.A., Adams L.G., Baumber A.J.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DE Distribution of bcf operon in Salmonella enterica.*;
 DE EMBL; AF130422; AAD34372.1;
 DE EMBL; AF129435; AAD22098.1;
 DE INTERPRO: IPR000015;
 DE PFAM: PF00577; Usher; 1.
 DR PROSITE: PS01151; FIMBRIAL_USHER; 1.
 SQ SEQUENCE 870 AA: 94667 MW: 9E6E7120B507B871 CRC64;
 Query Match 45.88; Score 44; DB 2; Length 870;
 Best Local Similarity 72.78; Pred. No. 3.3e+02;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 11 ALAVGDSNIG 21
 DB 418 ALAIGSGANIG 428
 RESULT 13
 ID C51918 PRELIMINARY; PRT: 992 AA.
 AC C51918
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE PROTEIN L PRECURSOR.
 OS Peptostreptococcus magnus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Peptostreptococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-3316;
 RA MEDLINE-95078460; PubMed-7987012;
 RA Murphy J.P., Townsends S.M., Ficht T.A., Adams L.G., Baumber A.J.;
 RT Nucleotide sequence of the gene for peptostreptococcal protein L.*;
 RL DNA Seq. 4:259-265(1994);
 DR EMBL; L04466; AAA67503.1;
 DR HSSP; O51911; IGAB.
 DR INTERPRO: IPR001899;
 DR INTERPRO: IPR003147;
 DR PFAM; PF0746; Gram_pos_anchor; 1.
 DR PFAM; PF0166; GN; 4;
 DR PFAM; PF02246; BI; 4;
 KW SIGNAL
 FT SIGNAL 1 24
 FT CHAIN 25 992
 FT PROTEIN L.
 SQ SEQUENCE 992 AA: 108699 MW: 9CFF5771578A5DCE CRC64;
 Query Match 45.88; Score 44; DB 2; Length 992;
 Best Local Similarity 50.08; Pred. No. 3.7e+02;
 Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
 QY 2 RVLVLAVALAVGDSN 19
 DB 6 KILWALLAGAVVGGAN 23
 RESULT 14
 ID Q9U280 PRELIMINARY; PRT: 1108 AA.
 AC Q9U280
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE Y49E10.11 PROTEIN.
 OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
 OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Barlow K.; Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

RL EMBL: Z58866; CAB11550.3; -

DR EMBL: Z58866; CAB11550.3; -

DR INTERPRO: IP0001757; -

DR PRINTS: PR00119; CATAPASE

DR PROSITE: PS00154; ATPASE_E1_E2; UNKNOWN.1

DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN.1

SQ SEQUENCE 1108 AA; 124865 MW; 298F1CD744799E2A CRC64;

Query Match 45.8%; Score 44; DB 5; Length 1108;

Best Local Similarity 64.3%; Pred. No. 4.2e+02;

Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 6 LALAVANGDSN 19

DB 755 LAKHVLAIIGDGN 768

RESULT 15

Q9V6P5

ID Q9V6P5 PRELIMINARY; PRT: 1297 AA.

AC Q9V6P5

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)

DE CGI7034 PROTEIN.

GN CGI7034

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;

RA MEDLINE-20196006; PubMed-10731132;

RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gossyne J.D.,

RA Annatides P.G., Scher S.E., Li W., Hoskins R.A., Galle R.P.,

RA Cheng K.A., Kuo S.-C., Ashburner M., Anderson S.M.,

RA Sutter G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Balow R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Chin J.J., Evangelista C., Ferraz C., Fierro S., Friesmann M.,

RA Foster C., Gabor C., Garg R., Gellet M., Gellert M., Harris K.,

RA Glodak A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,

RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merklov G., Milshina N.Y., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman C.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,
 RA Zheng X.H., Zhong F.N., Zhong M., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA The genome sequence of *Drosophila melanogaster*;

RL Sequence Z58866; CAB11550.3; -

DR EMBL: Z58866; CAB11550.3; -

DR HSP: Q41261; 100U

DR FLYBASE: FBgn0033837; CGI7034.

DR INTERPRO: IP0001757; -

DR PFAM: PF00122; E1-E2 ATPase; 3.

DR PRINTS: PR00119; CATAPASE.

DR PROSITE: PS00154; ATPASE_E1_E2; 1.

SQ SEQUENCE 1297 AA; 147293 MW; B11ED6E75B4D9D8A CRC64;

Query Match 45.8%; Score 44; DB 5; Length 1297;

Best Local Similarity 72.7%; Pred. No. 4.9e+02;

Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 9 AVALAVGDSN 19

DB 843 AVTLAIGDGN 853

Search completed: May 3, 2001, 15:52:43

Job time: 62 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2001, 21:23:56 ; Search time 1035.58 Seconds
(without alignments)
1139.293 Million cell updates/sec

Title: US-09-426-776-11
Perfect score: 80
Sequence: 1 attacatccaccagccatg.....gggaccagtcacacttggg 80

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_bal:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pil:*
- 13: gb_p12:*
- 14: gb_p13:*
- 15: gb_p14:*
- 16: em_bal:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_vi:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_vil:*

59: gb_v12:*

60: gb_htg1:*

61: gb_htg2:*

62: gb_htg3:*

63: gb_htg4:*

64: gb_htg5:*

65: gb_htg6:*

66: gb_htg7:*

67: gb_htg8:*

68: gb_htg9:*

69: gb_htg10:*

70: gb_htg11:*

71: gb_htg12:*

72: gb_htg13:*

73: gb_htg14:*

74: gb_htg15:*

75: gb_htg16:*

76: gb_htg17:*

77: gb_htg18:*

78: gb_htg19:*

79: gb_htg20:*

80: gb_htg21:*

81: gb_htg22:*

82: gb_htg23:*

83: gb_htg24:*

84: gb_htg25:*

85: gb_pr1:*

86: gb_pr2:*

87: gb_pr3:*

88: gb_pr4:*

89: gb_pr5:*

90: gb_pr6:*

91: gb_pr7:*

92: gb_pr8:*

93: gb_pr9:*

94: gb_ro1:*

95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	74	92.5	5511	8	AF017250
2	55.8	69.8	11453	8	AF072686 Oreochrom
3	43.4	54.2	5166	8	U70826 Fundulus he
c	29	36.2	42805	85	AC004221 Homo sapi
5	28.8	36.0	945	92	AL137301 Homo sapi
6	28.8	36.0	4214	89	AK023132 Homo sapi
7	28.8	36.0	4813	89	AK024508 Homo sapi
8	28.8	36.0	139505	92	HSJ591C20
c	28.6	35.8	153713	74	AC068717 Homo sapi
10	28.4	35.5	10807	9	AR070469
11	28.4	35.5	10807	10	I47708 Sequence 7

12	28.4	35.5	172815	62	AC010920	AC010920 Drosophil
C 13	28.2	35.2	840	6	S66686	S66686 n-syb=vamp
C 14	28.2	35.2	50701	65	AC018231	AC018231 Drosophil
C 15	28.2	35.2	146271	3	SYCSLRB	D64000 Synecocyst
C 16	28.2	35.2	257867	4	AC005557	AC005557 Drosophil
17	28.2	35.2	303191	4	AE003472	AE003472 Drosophil
18	27.6	34.5	272	7	OALGLR	W7004 10.aries ge
19	27.6	34.5	785	7	OALGLR	X04520 Ovine mRNA
20	27.6	34.5	824	9	AR002668	AR002668 Sequence
21	27.6	34.5	824	10	I43695	I43695 Sequence
22	27.6	34.5	1769	7	SHPBLGAL	M32233 Ovis beta-1
23	27.6	34.5	7379	7	OALGB	X12817 Ovis aries
C 24	27.6	34.5	62838	64	AC015633	AC015633 Homo sapi
C 25	27.2	34.2	170245	91	HS109FI14	AL022721 Human DNA
C 26	27.2	34.0	190289	59	MCU60315	U60315 Molluscum c
C 27	27.2	34.0	211123	81	AL450163	AL450163 Homo sapi
C 28	27	33.8	144096	80	AL359312	AL359312 Homo sapi
C 29	27	33.8	144981	82	AL513014	AL513014 Mus muscu
C 30	27	33.8	152044	73	AC068235	AC068235 Homo sapi
31	27	33.8	152211	7	S5C251829	AJ251829 Sus scrof
32	26.8	33.5	902	56	XXU49832	U49832 Synthetic s
33	26.8	33.5	4186	9	AX026418	AX026418 Sequence
34	26.8	33.5	4187	9	AX026417	AX026417 Sequence
35	26.8	33.5	4187	93	HUMRSC390	DI3643 Homo sapien
36	26.8	33.5	4248	9	AX026416	AX026416 Sequence
37	26.8	33.5	31568	87	AF261758	AF261758 Homo sapi
38	26.8	33.5	151802	87	AC020956	AC020956 Homo sapi
C 39	26.8	33.5	151802	79	AL354867	AL354867 Homo sapi
C 40	26.8	33.5	163479	62	AC010616	AC010616 Homo sapi
C 41	26.8	33.5	169072	87	AC009946	AC009946 Homo sapi
C 42	26.6	33.2	2248	14	PAB302017	AC009946 Homo sapi
C 43	26.6	33.2	242328	74	AC069308	AJ302017 Picea abi
C 44	26.4	33.0	3708	52	HS69A	AC069308 Mus muscu
C 45	26.4	33.0	7791	59	SOUCAPEPRO	X69838 H. sapiens m
			330	92		L07418 Southampton

ALIGNMENTS

```

RESULT      1
AF017250
LOCUS      5511 bp      mRNA
DEFINITION Oreochromis aureus vitellogenin precursor (Vtgl) mRNA, complete cds.
ACCESSION AF017250
VERSION   AF017250.1
KEYWORDS  GI:4102880
SOURCE    Oreochromis aureus
ORGANISM  Oreochromis aureus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes;
Labroidae; Cichlidae; Oreochromis.
1 (bases 1 to 5511)
Lim,E.H., Lam,T.J. and Ding,J.L.
Cloning of Full-length Oreochromis aureus Vitellogenin cDNA and its
Deduced Primary Structure
Unpublished
2 (bases 1 to 5511)
Lim,E.H., Lam,T.J. and Ding,J.L.
Direct Submission
Submitted (05-AUG-1997) School of Biological Sciences, National
University of Singapore, 10 Kent Ridge Crescent 119260, Singapore
Location/Qualifiers
1. .5511
/organism="Oreochromis aureus"
/db_xref="taxon:47969"
/sex="female"
/tissue_type="liver"
/dev_stage="estradiol-induced adult"
/clone="pOavtgl"
1. .13
/genes="Vtgl"
5'UTR

```

gene

[illegible]

RESULT 2
AF072686
LOCUS

18-OCT-1996
RNA, complete cds.


```
gene          /evidence=not_experimental
2489..3898
/ gene="dJ591C20.1"
CDS           / gene="dJ591C20.1"
2489..3898
/ note="match: cDNAs: Em:AK001207 Em:AL157426 Em:U95607
Em:X57435
match: ESTs: Em:AA664327 Em:AI207810 Em:AA779058
match: proteins: Tr:O95870 Tr:O92102 Tr:O69340 Tr:O9VZF1
Sw:P41879 Tr:O69130 Tr:O90783 Sw:P04258"
/codon_start=1
/evidence=not_experimental
/product="dJ591C20.1 (novel protein similar to mouse
K626)"
/protein_id="CAC15491.1"
/db_xref="GI:11125672"
/translation="MCVICFKALVRVFKIYLTASYTPFRGWPVAFRWDVAVRGRS
SSHRALTCAAAAGVLLRDETGLDGRPPRGARSOACLLQQLRELPGOLASVAL
AHSGLRWLVPGSVSLMTRALLPLLOQGOERLVERYHGRKLVACDGNEDTMEFMDR
ROHPSGVHGPLVITCCGNAGFYEMGCLSPLEAGYSVLGNHHPFGSSGTVPFPQH
DANAMDVVEYALHRLHFPFAHVLVYGVSGVGTATWATYPELGAFLVDATFDLVL
PLAKMVPHSWKGVLVYTRVREHFNLVAEQLCYPCPVILLRLRTQDDVYSTGRPLPL
SPGDVGNKGNELLRLLEHRYPVVMAREGRAVTRWLRAGSLAQAFAFYARYVDEED
WCLALLRSYRACRESELEGEALGPHGPAFPLVQOGLSSRRRLALFLARKHLKNV
EATHSPLEPEEFQPLRWL"
5103..5232
/ note="MIR repeat: matches 9..146 of consensus"
repeat_region / note="AluSg repeat: matches 6..312 of consensus"
repeat_region / join(6238..6332,10244..10389,14616..14764,16764..16823,
mRNA          23667..23768,30138..30186,30798..32493)
/ gene="TPD52L2"
/ note="match: cDNAs: Em:AF004430
match: ESTs: Em:AW230830 Em:AA833100 Em:AA608309"
/evidence=not_experimental
/product="dJ591C20.2.1 (tumor protein D52-like 2)"
6238..32493
/ gene="TPD52L2"
Join(6314..6332,10244..10389,14616..14764,16764..16823,
CDS           23667..23768,30138..30186,30798..30893)
/ gene="TPD52L2"
/ note="isoform 1
match: proteins: Tr:O43399"
/codon_start=1
/evidence=not_experimental
/product="dJ591C20.2.1 (tumor protein D52-like 2)"
/protein_id="CAC15492.1"
/db_xref="GI:11125673"
/translation="MDSAGQDNLNPNKGLLSDMTDVPVDTGVAARTPAVEGLTEA
EEELRLAETKVEEIVTLRQVLAAKERHCGLKRLGLSTGLKQNLRSWHDVQV
SSAYVKTSEKLGEMNEKVQSDLYKRTQETLSQAGKTSAAALSTVGSATSRKLGDMRN
SATFKSFEDRVGTIKSVKVVGDRENGSDNLPSSAGSGDKPLSDPAPE"
7042..7350
/ note="AluSc repeat: matches 1..309 of consensus"
repeat_region / note="AluSg repeat: matches 1..309 of consensus"
repeat_region / note="AluY repeat: matches 1..309 of consensus"
repeat_region / note="AluJo repeat: matches 134..311 of consensus"
repeat_region / note="AluSp repeat: matches 1..308 of consensus"
repeat_region / note="AluJo repeat: matches 120..134 of consensus"
repeat_region / note="AluSx repeat: matches 1..300 of consensus"
repeat_region / note="AluSg/x repeat: matches 1..117 of consensus"
repeat_region / note="AluSx repeat: matches 2..212 of consensus"
repeat_region / note="AluSx repeat: matches 5..292 of consensus"
misc_feature  complement(971..10250)
misc_feature  / note="match: GSS: Em:AQ888148"
complement(9810..10258)
```

```
repeat_region / note="match: GSS: Em:AQ583076"
10791..11090
/ note="AluDb repeat: matches 1..298 of consensus"
repeat_region / note="AluY repeat: matches 106..306 of consensus"
11469..11666
/ note="AluSg repeat: matches 1..297 of consensus"
11667..11963
/ note="AluSg repeat: matches 1..297 of consensus"
11989..12290
/ note="AluSx repeat: matches 1..302 of consensus"
repeat_region / note="AluSx repeat: matches 21..295 of consensus"
12302..12577
/ note="AluSg repeat: matches 1..298 of consensus"
12587..12881
/ note="AluSg repeat: matches 1..298 of consensus"
complement(12730..13211)
misc_feature / note="match: GSS: Em:AQ192102"
12986..13011
/ note="match: GSS: Em:AQ192102"
13231..13511
/ gene="TPD52L2"
/ note="match: GSS: Em:AQ722319"
13322..13625
/ note="AluSx repeat: matches 1..306 of consensus"
13639..13960
/ note="AluY repeat: matches 1..311 of consensus"
14156..14464
/ note="AluSp repeat: matches 1..309 of consensus"
14989..15297
/ note="AluSx repeat: matches 31..311 of consensus"
15312..15622
/ note="AluJo repeat: matches 1..312 of consensus"
15806..16101
/ note="AluSx repeat: matches 1..294 of consensus"
16102..16127
/ note="match: GSS: Em:AQ631401"
17964..18272
/ note="AluY repeat: matches 1..309 of consensus"
18273..18406
/ note="AluY repeat: matches 1..137 of consensus"
18776..18979
/ note="match: GSS: Em:AQ631401"
19859..19957
/ note="MIR repeat: matches 119..225 of consensus"
21557..21857
/ note="AluY repeat: matches 1..300 of consensus"
21557..21769
/ gene="TPD52L2"
/ note="match: GSS: Em:AQ631401"
22007..22329
/ gene="TPD52L2"
Query Match 36.0%; Score 28.8; DB 92; Length 139505;
Best Local Similarity 60.0%; Pred. No. 17;
Matches 48; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
```

```
QY 1 attcacatccaccagcatgagggtgcttgctactagctctgtgtgctcgcagtg99 60
|||||
Db 76746 AATCAATGACACGATGGAAGTCTGTCTCATGCTGTTTGTCTACTGCGCTCT 76805
|||||
QY 61 gggaccagtccaacttgggg 80
|||||
Db 76806 GGAATCCGACCGTGTGGGG 76825
|||||
RESULT 9
AC068717/c 153713 bp DNA HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 4 clone RP11-13K9, WORKING DRAFT SEQUENCE,
DEFINITION 13 unordered pieces.
AC068717
AC068717.2 GI:8468941
VERSION AC068717.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2001, 15:51:41 ; Search time 12.84 Seconds
(without alignments)
112.397 Million cell updates/sec

Title: US-09-426-776-10
Perfect score: 96
Sequence: 1 MRVLVALAVALAVGDSNLG 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues
Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_67:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	85.4	1788	2 T31095	vitellogenin precu
2	53	55.2	678	2 S46963	exed protein - Aer
3	53	55.2	1687	2 T43144	vitellogenin II pr
4	45	46.9	330	2 T25169	hypothetical prote
5	45	46.9	691	2 T41724	probable adenosine
6	44	45.8	328	2 H75251	glycerol-3-phospha
7	44	45.8	585	2 S06938	sphingomyelin phos
8	44	45.8	629	1 A39825	sphingomyelin phos
9	44	45.8	992	2 S54396	protein L precursor
10	43	44.8	127	2 B75353	hypothetical prote
11	43	44.8	222	2 JC1384	beta-casein precu
12	43	44.8	273	2 C70551	hypothetical prote
13	43	44.8	346	2 H82637	glycerol-3-phospha
14	43	44.8	359	2 T44816	brp protein (impor
15	43	44.8	583	2 H75561	ABC transporter, A
16	43	44.8	598	2 T42070	protein serine/thr
17	43	44.8	683	2 I52996	transforming growt
18	43	44.8	729	2 T46328	probable adenosine
19	42.5	44.3	80	2 A65210	hypothetical 7.4 k
20	42.5	44.3	310	2 C83125	hypothetical prote
21	42.5	44.3	893	2 T03864	hypothetical prote
22	42	43.8	168	2 A31770	pheromone-binding
23	42	43.8	174	2 A64442	hypothetical prote
24	42	43.8	258	2 JH0472	apolipoprotein A-I
25	42	43.8	297	2 D69404	hypothetical prote
26	42	43.8	378	2 D64181	probable cytochrom
27	42	43.8	460	2 C70576	probable mgE prot
28	42	43.8	550	2 C64867	probable sulfate t
29	42	43.8	654	2 H65125	probable general s

30	41.5	43.2	351	2 B81721	cytochrome D ubiqu
31	41	42.7	109	2 D75506	hypothetical prote
32	41	42.7	180	1 LGST	beta-lactoglobulin
33	41	42.7	180	1 LGST	beta-lactoglobulin
34	41	42.7	222	2 A32979	beta-casein precu
35	41	42.7	224	1 KB0A2	beta-casein precu
36	41	42.7	226	1 KBHU	beta-casein precu
37	41	42.7	228	2 J70564	beta-casein precu
38	41	42.7	231	2 F69147	UDP-N-acetylglucos
39	41	42.7	242	2 G82956	hypothetical prote
40	41	42.7	312	2 S18387	heme oxygenase - r
41	41	42.7	345	2 S55377	urPAB protein prec
42	41	42.7	387	2 A53586	albumin-binding pr
43	41	42.7	397	2 B70763	probable membrane
44	41	42.7	429	2 C83025	probable phosphose
45	41	42.7	597	2 C69283	hypothetical prote

ALIGNMENTS

RESULT 1
T31095
vitellogenin precursor - Oreochromis aureus
C:Species: Oreochromis aureus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
C:Accession: T31095
R:Lim, E.H.; Lam, T.J.; Ding, J.L.
Submitted to: Cloning of full-length Oreochromis aureus vitellogenin cDNA and its
A:Description: Cloning of full-length Oreochromis aureus vitellogenin cDNA and its
A:Reference number: Z20978
A:Accession: T31095
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1788 <LIM>
A:Cross-references: EMBL:AF017250; NID:g4102880; PID:g4102881; PIDN:AAD01615.1
C:Superfamily: vitellogenin

Query Match 85.4%; Score 82; DB 2; Length 1788;
Best Local Similarity 95.0%; Pred. No. 0.0062;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRVLVALAVALAVGDSNL 20
DB 1 MRVLVALAVALAVGDSNL 20
|||||

RESULT 2
S46963
exed protein - Aeromonas salmonicida
C:Species: Aeromonas salmonicida
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C:Accession: I39678; S46963
R:Kariyshev, A.V.; MacIntyre, S.
Gene 158, 77-82, 1995
A:Title: Cloning and study of the genetic organization of the exe gene cluster of
A:Reference number: I39675; MUID:95309729
A:Accession: I39678
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-678 <RES>
A:Cross-references: EMBL:X80505; NID:g516349; PIDN:CAA56668.1; PID:g516351
C:Genetics:
A:Gene: exed

Query Match 55.2%; Score 53; DB 2; Length 678;
Best Local Similarity 60.0%; Pred. No. 4.1;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 RVLVALAVALAVGDSNLG 21
: ||| | : : ||| |||

Db 352 QVLVEALIVEIADGDLNLG 371

RESULT 3

T43144

C:Species: Fundulus heteroclitus (mummichog)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C:Accession: T43144
R:Laflaur, G.J.; Hoch, K.L.; Denslow, N.; Byrne, B.M.; Wallace, R.A.
submitted to the EMBL Data Library, September 1996
A:Description: Derivation of oocyte and egg proteins from parental vitellogenins in Fundulus heteroclitus
n.
A:Reference number: 22316
A:Accession: T43144
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1687 <LAF>
A:Cross-references: EMBL:U70826; NID:91621358; PID:91621359; PIDN:AAB17152.1
A:Experimental source: estrogen-induced liver
C:Superfamily: vitellogenin
C:Keywords: egg yolk; phosphoprotein
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-1687/Product: vitellogenin II #status predicted <MAT>

Query Match 55.2%; Score 53; DB 2; Length 1687;
Best Local Similarity 75.0%; Pred. No. 8.7;
Matches 12; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MRVLVALAVAVCD 16

Db 1 MRVLVALAVAVCD 16

RESULT 4

T25169

hypothetical protein T23F1.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T25169
R:Wilkinson, J.
submitted to the EMBL Data Library, October 1996
A:Reference number: 21990
A:Accession: T25169
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-330 <WIL>
A:Cross-references: EMBL:Z81129; PIDN:CAB03405.1; GSPDB:GN00023; CESP:T23F1.6
A:Experimental source: clone T23F1
C:Genetics:
A:Gene: CESP:T23F1.6
A:Map position: 5
A:Introns: 16/3
C:Superfamily: gliadin

Query Match 46.9%; Score 45; DB 2; Length 330;
Best Local Similarity 45.0%; Pred. No. 31;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 MRVLVALAVAVCDGSL 20

Db 1 MRVLVALAVAVCDGSL 20

RESULT 5

T41724

probable adenosinetriphosphatase (EC 3.6.1.3) SPAC821.13c [similarity] - fission yeast
C:Species: Schizosaccharomyces pombe
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000
C:Accession: T41724
R:Rieger, M.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, October 1999

A:Reference number: 22012
A:Accession: T41724
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-691 <RIE>
A:Cross-references: EMBL:AL121770; PIDN:CAB57447.1; GSPDB:GN00066; SPDB:SPAC821.13c
A:Experimental source: strain 972h-; cosmid c821
C:Genetics:
A:Gene: SPDB:SPAC821.13c
A:Map position: 1
C:Keywords: hydrolase

Query Match 46.9%; Score 45; DB 2; Length 691;
Best Local Similarity 64.3%; Pred. No. 58;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 6 LALVALAVGDSN 19

Db 362 LEKAVTLAIGDGN 375

RESULT 6

H75251

glycerol-3-phosphate dehydrogenase, NAD(+)-dependent - Deinococcus radiodurans (str.)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: H75251
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.A.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: H75251
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-328 <WHI>
A:Cross-references: GB:AE002091; GB:AE000513; NID:96460446; PIDN:AAF12158.1; PID:96460446
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2621
A:Map position: 1
C:Superfamily: glycerol-3-phosphate dehydrogenase

Query Match 45.8%; Score 44; DB 2; Length 328;
Best Local Similarity 52.6%; Pred. No. 44;
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 VLVLALAVAVGDSN 21

Db 184 VLKNIIVAVAGMGDLHLG 202

RESULT 7

S06958

sphingomyelin phosphodiesterase, acidic, inactive splice form 2 - human
N:Alternate names: acid sphingomyelinase, splice form 2
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1993 #sequence_revision 24-Nov-1999 #text_change 21-Jan-2000
C:Accession: S06958; A39825
R:Quintern, L.E.; Schuchman, E.H.; Levrin, O.; Suchi, M.; Ferlinz, K.; Reinke, H.; S. EMBO J. 8, 2469-2473, 1989
A:Title: Isolation of cDNA clones encoding human acid sphingomyelinase: occurrence of
A:Reference number: S06957; MUID:90060003
A:Accession: S06958
A:Molecule type: mRNA
A:Residues: 270-585 <QUI>
A:Cross-references: EMBL:X52679; NID:928881; PIDN:CAA36902.1; PID:9825629
R:Schuchman, E.H.; Suchi, M.; Takahashi, T.; Sandhoff, K.; Desnick, R.J.
J. Biol. Chem. 266, 8531-8539, 1991

A:Title: Human acid sphingomyelinase. Isolation, nucleotide sequence, and expression of
A:Reference number: A39825; MUID:91217097
A:Accession: A39825
A:Molecule type: mRNA
A:Residues: 1-362, 'TGGFVALSPYGLRLISLNMFCRSRENEWL', 'INSTDPAGQLQWLVLGELQAAEDRGD', 375-585
A:Cross-references: GB:M59916; NID:g179094; PIDN:AAA58377.1; PID:g179095
A:Note: parts of this sequence were confirmed by peptide sequencing
R:Levran, O.; Desnick, R.J.; Schuchman, E.H.
J. Biol. Invest. 88, 806-810, 1991
C:Comment: This form lacked catalytic activity when expressed in COS-1 cells. Its function
C:Comment: For the active splice form 1, see PIR:A39825.
C:Genetics:
A:Gene: GDB:SMPD1
A:Cross-references: GDB:128144; OMIM:257200
A:Map position: 11p15.4-11p15.4
A:Introns: 104/3; 362/2; 375/3; 401/2; 450/1
C:Superfamily: acid sphingomyelinase; phosphoesterase core homology; saposin repeat homology
C:Keywords: alternative splicing
F:81-175/Domain: saposin repeat homology <SAP>
F:200-320/Domain: phosphoesterase core homology <PEC>

Query Match 45.8%; Score 44; DB 2; Length 585;
Best Local Similarity 71.4%; Pred. No. 71;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VLVLALVALAVGD 16
|| |||||:|:
Db 36 VLALALALALSD 49

RESULT 8
A39825
sphingomyelin phosphodiesterase (EC 3.1.4.12), acidic, splice form 1 precursor - human
N:Alternate names: acid sphingomyelinase
C:Species: Homo sapiens (man)
C:Date: 20-Mar-1992 #sequence_revision 07-Jul-1995 #text_change 24-Nov-1999
C:Accession: S27009; S36357; A39825; JX0276; S06957; I55567; A42801; B42801; D42801
R:Newrzel, D.; Stoffel, W.
Biol. Chem. Hoppe-Seyler 373, 1233-1238, 1992
A:Title: Molecular cloning of the acid sphingomyelinase of the mouse and the organization
A:Reference number: S27009; MUID:93183402
A:Accession: S27009
A:Molecule type: DNA
A:Residues: 1-629 <NEW>
A:Cross-references: EMBL:X63600; NID:g556808
R:Hoimann, K.
submitted to the EMBL Data Library, December 1991
A:Reference number: S36357
A:Accession: S36357
A:Molecule type: DNA
A:Residues: 1-321, 'T', 323-629 <HOF>
A:Cross-references: EMBL:X63600; NID:g556808; PIDN:CAA45145.1; PID:g556809
R:Schuchman, E.H.; Suchi, M.; Takahashi, T.; Sandhoff, K.; Desnick, R.J.
J. Biol. Chem. 266, 8531-8539, 1991
A:Title: Human acid sphingomyelinase. Isolation, nucleotide sequence, and expression of
A:Reference number: A39825; MUID:91217097
A:Accession: A39825
A:Molecule type: mRNA
A:Residues: 1-629 <SCH>
A:Cross-references: GB:M59916; NID:g179094; PIDN:AAA58377.1; PID:g179095
A:Note: polymorphisms were demonstrated at positions 322 and 506
R:Ida, H.; Rennett, O.M.; Eto, Y.; Chan, W.Y.
J. Biochem. 114, 15-20, 1993
A:Title: Cloning of a human acid sphingomyelinase cDNA with a new mutation that renders
A:Reference number: JX0276; MUID:94012573
A:Accession: JX0276
A:Molecule type: mRNA
A:Residues: 1-35, 38-156, 'R', 158-321, 'T', 323-505, 'G', 507-629 <IDA>
A:Cross-references: EMBL:X59960; NID:g402620; PIDN:CAA42584.1; PID:g402621
A:Note: the authors translated the codon CTC for residue 85 as Thr and ACC for residue 8
A:Note: this form has no catalytic activity; correction of position 157 to Cys restores
R:Quinter, L.E.; Schuchman, E.H.; Levran, O.; Suchi, M.; Ferlinz, K.; Reinke, H.; Sandhoff, J.
EMBO J. 8, 2469-2473, 1989
A:Title: Isolation of cDNA clones encoding human acid sphingomyelinase: occurrence of al
A:Reference number: S06957; MUID:90060003

A:Accession: S06957
A:Molecule type: mRNA
A:Residues: 128-629 <QUT>
A:Cross-references: EMBL:X52678; NID:g28879; PIDN:CAA36901.1; PID:g28880
A:Note: parts of this sequence were confirmed by peptide sequencing
R:Levran, O.; Desnick, R.J.; Schuchman, E.H.
J. Clin. Invest. 88, 806-810, 1991
A:Title: Niemann-Pick type B disease. Identification of a single codon deletion in
A:Reference number: I55567; MUID:91358737
A:Accession: I55567
A:Status: translation not shown; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 119-120, 'H', 122 <LEW>
A:Cross-references: GB:S55766; NID:g234719; PIDN:AA19680.1; PID:g234720
R:Takahashi, T.; Suchi, M.; Desnick, R.J.; Takada, G.; Schuchman, E.H.
J. Biol. Chem. 267, 12552-12558, 1992
A:Title: Identification and expression of five mutations in the human acid sphingomyelinase and non-neuropathic forms.
A:Reference number: A42801; MUID:92316934
A:Contents: annotation; characterization of mutations
A:Note: substitution of Ile for 382-Met or Ser for 383-Asn result in complete inactivation and lead to milder (type B) disease
C:Comment: Two isoforms, neutral and acidic, have been identified. The acidic isoform is the inactive splice form 2, see PIR:S06958.
C:Genetics:
A:Gene: GDB:SMPD1
A:Cross-references: GDB:128144; OMIM:257200
A:Map position: 11p15.4-11p15.4
A:Introns: 104/3; 362/2; 419/3; 445/2; 494/1
A:Note: a defect in this gene may result in Niemann-Pick disease
C:Complex: monomer
C:Function:
A:Description: catalyzes the hydrolysis of sphingomyelin to form phosphorylcholine
C:Superfamily: acid sphingomyelinase; phosphoesterase core homology; saposin repeat homology
C:Keywords: acetylated amino end; alternative splicing; glycoprotein; lysosome; membrane
F:1-48/Domain: signal sequence #status predicted <SIG>
F:49-629/Product: sphingomyelin phosphodiesterase #status predicted <MAT>
F:81-175/Domain: saposin repeat homology <SAP>
F:200-320/Domain: phosphoesterase core homology <PEC>
F:49/Modified site: acetylated amino end (Asp) (in mature form) #status predicted
F:86,175,335,395,503,520/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:89-165,92-157,120-131/Disulfide bonds: #status predicted
F:251,319,439/Active site: Asp, His, Tyr #status predicted

Query Match 45.8%; Score 44; DB 1; Length 629;
Best Local Similarity 71.4%; Pred. No. 75;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VLVLALVALAVGD 16
|| |||||:|:
Db 36 VLALALALALSD 49

RESULT 9
S54396
protein L precursor - Peptostreptococcus magnus (strain 3316)
C:Species: Peptostreptococcus magnus
A:Variety: strain 3316
C:Date: 27-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 15-Oct-1999
C:Accession: S54396
R:Murphy, J.P.; Duggleby, C.J.; Atkinson, M.A.; Town, A.R.; Atkinson, T.; Gowa
Mol. Microbiol. 12, 911-920, 1994
A:Title: The functional units of a peptostreptococcal protein L.
A:Reference number: S54396; MUID:95020613
A:Accession: S54396
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-992 <MUR>
A:Cross-references: EMBL:L04466; NID:g150673; PIDN:AAA67503.1; PID:g150674

Query Match 45.8%; Score 44; DB 2; Length 992;

Best Local Similarity 50.0%; Pred. No. 11e+02;
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 RVLVLALAVAGDGSN 19
DB 6 KLLMALAGAIIVGGGAN 23

RESULT 10

B75353

hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 26-May-2000

C:Accession: B75353

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M.S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: B75353

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-127 <WHI>

A:Cross-references: GB:AE0002021; GB:AE000513; MID:g6459573; PIDN:AAFL1348.1; PID:g645956

A:Experimental source: strain R1

C:Genetics:

A:Gene: DRI788

A:Map position: 1

C:Superfamily: Deinococcus radiodurans hypothetical protein DRI788

Query Match 44.8%; Score 43; DB 2; Length 127;
Best Local Similarity 69.2%; Pred. No. 28;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRVLVLAVALA 13
DB 1 MRILLVALAALA 13

RESULT 11

JC1384

beta-casein precursor - goat

C:Species: Capra aegagrus hircus (domestic goat)

C>Date: 10-Jun-1993 #sequence_revision 10-Jun-1993 #text_change 23-Feb-1997

C:Accession: JC1384

R:Roberts, B.; DiTullio, P.; Vitale, J.; Hehir, K.; Gordon, K.

Gene 121, 255-262, 1992

A:Title: Cloning of the goat beta-casein-encoding gene and expression in transgenic mice

A:Reference number: JC1384; MUID:93077039

A:Accession: JC1384

A:Molecule type: DNA

A:Residues: 1-222 <ROB>

A:Cross-references: GB:M90556

C:Genetics:

A:Gene: CSN2

A:Introns: 17/3; 26/3; 35/3; 45/3; 57/3; 221/3

A:Superfamily: beta-casein

C:Keywords: milk; phosphoprotein

Query Match 44.8%; Score 43; DB 2; Length 222;
Best Local Similarity 64.3%; Pred. No. 44;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MRVLVLAVALAV 14
DB 1 MKVLILACLVALAI 14

RESULT 12

C70551

hypothetical protein RV2507 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70551

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: C70551

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-273 <COL>

A:Cross-references: GB:295556; GB:AL123456; MID:g3561778; PIDN:CAB08925.1; PID:e3161

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV2507

Query Match 44.8%; Score 43; DB 2; Length 273;
Best Local Similarity 50.0%; Pred. No. 52;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 3 VLVLALAVAGDGS 18
DB 120 LLVVALVALVANGS 135

RESULT 13

H82637

glycerol-3-phosphate dehydrogenase XF1802 [imported] - Xylella fastidiosa (strain 9)

C:Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000

C:Accession: H82637

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Se

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: H82637

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-346 <SIM>

A:Cross-references: GB:AE004001; GB:AE003849; MID:g9106864; PIDN:AAF84610.1; GSPDB:C

A:Experimental source: strain 9a5C

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, I

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; I

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; I

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Marti

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, R.C.; Miyaki, C

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.G.; Palmieri,

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sa

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Sil

M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.I

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF1802

C:Superfamily: glycerol-3-phosphate dehydrogenase

Query Match 44.8%; Score 43; DB 2; Length 346;
Best Local Similarity 64.3%; Pred. No. 63;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 8 LVALAVDGSNGLG 21
DB 197 LAVAIGVADGMQLG 210

RESULT 14

T44816
brp protein [imported] - Halobacterium salinarum
C:Species: Halobacterium salinarum
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 31-Jan-2000
C:Accession: T44816
R:Betlach, M.; Friedman, J.; Boyer, H.W.; Pfeifer, F.
Nucleic Acids Res. 12, 7949-7959, 1984
A:Title: Characterization of a halobacterial gene affecting bacterio-opsin gene expression
A:Reference number: 222850; MUID:85037955
A:Accession: T44816
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-359 <BET>
A:Cross-references: EMBL:X01081; PIDN:CAA25558.1
A:Note: the source is designated as Halobacterium halobium
C:Genetics:
A:Gene: brp

Query Match 44.8%; Score 43; DB 2; Length 359;
Best Local Similarity 58.8%; Pred. No. 65;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 RVLVLALAVAVGDS 18
I: |||| | || I:
Db 26 RLSELLAAFAVGYGA 42

RESULT 15

H75561
ABC transporter, ATP-binding protein, MsbA family - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: H75561
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Otterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: H75561
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-583 <WHI>
A:Cross-references: GB:AE001872; GB:AE000513; NID:g6457750; PIDN:AAF09688.1; PID:g645776
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0096
A:Map position: 1
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 44.8%; Score 43; DB 2; Length 583;
Best Local Similarity 55.0%; Pred. No. 98;
Matches 11; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 1 MRVLVLALAVAVGDSNL 20
I: | || | || | I:
Db 18 LRGAVALAALLLVGTGLNL 37

Search completed: May 3, 2001, 15:53:11
Job time: 90 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2001, 21:25:01 ; Search time 65.18 Seconds
(without alignments)
214,302 Million cell updates/sec

Title: US-09-426-776-11
Perfect score: 80
Sequence: 1 attaccatccaccagccatg.....gggaccagtccaactgggg 80

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28.4	35.5	10807	1	US-08-206-176-7
2	28.4	35.5	10807	2	US-08-756-506-5
3	27.6	34.5	824	1	US-08-154-019-29
4	27.6	34.5	824	1	US-08-461-333-29
5	27.6	34.5	824	3	US-08-464-167-29
6	27.6	34.5	824	3	US-09-158-313-29
7	27.6	34.5	824	4	US-08-476-798-29
8	25.4	31.8	1607	3	US-08-753-007A-3
9	25.4	31.8	1607	4	US-09-398-496-3
10	25.4	31.8	2467	3	US-08-753-007A-1
11	25.4	31.8	2467	4	US-09-398-496-1
12	24.2	30.2	2296	2	US-08-899-336-2
13	23.8	29.8	2018	2	US-08-557-973-1
14	23.8	29.8	35081	2	US-08-752-760A-1
15	23.6	29.5	274	4	US-08-444-818-63
16	23.6	29.5	1788	2	US-08-722-808A-1
17	23.6	29.5	3495	1	US-08-446-038B-2
18	23.6	29.5	3495	1	US-08-446-010B-2
19	23.6	29.5	3495	1	US-08-805-445-2
20	23.6	29.5	3495	2	US-08-064-067D-2
21	23.6	29.5	3495	2	US-09-066-208-2
22	23.6	29.5	3629	1	US-08-097-997A-8
23	23.6	29.5	3629	4	US-08-665-574C-8
24	23.6	29.5	3629	4	US-08-946-994-8
25	23.6	29.5	6785	4	US-08-444-818-65
26	23.6	29.5	7310	4	US-08-444-818-74
27	23.6	29.5	8316	4	US-08-444-818-88

C 28	23.6	29.5	8987	4	US-08-444-818-137	Sequence 137, App
C 29	23.6	29.5	9185	4	US-08-444-818-122	Sequence 122, App
C 30	23.6	29.5	9185	4	US-08-444-818-123	Sequence 123, App
C 31	23.6	29.5	9379	4	US-08-444-818-176	Sequence 176, App
C 32	23.6	29.5	9401	1	US-07-910-760-9	Sequence 9, Appli
C 33	23.6	29.5	9401	1	US-08-440-519-9	Sequence 9, Appli
C 34	23.6	29.5	9401	5	PCT-US91-02225-9	Sequence 9, Appli
C 35	23	28.8	3311	2	US-08-239-276-10	Sequence 10, Appl
C 36	23	28.8	3311	2	US-08-468-579B-10	Sequence 10, Appl
C 37	23	28.8	3311	3	US-08-468-577B-10	Sequence 10, Appl
C 38	22.8	28.5	2458	3	US-08-756-806A-65	Sequence 65, Appl
C 39	22.8	28.5	2458	3	US-09-143-214-65	Sequence 65, Appl
C 40	22.8	28.5	2793	2	US-08-347-563A-1	Sequence 1, Appli
C 41	22.8	28.5	2793	3	US-08-485-942A-1	Sequence 1, Appli
C 42	22.8	28.5	2793	4	US-08-488-214A-1	Sequence 1, Appli
C 43	22.8	28.5	2793	4	US-08-488-208A-1	Sequence 1, Appli
C 44	22.8	28.5	12912	2	US-08-460-751-1	Sequence 1, Appli
C 45	22.8	28.5	14060	3	US-08-658-136-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-206-176-7
; Sequence 7, Application US/08206176
; Patent No. 5639940
; GENERAL INFORMATION:
; APPLICANT: Garner, Ian
; APPLICANT: Dairymple, Michael A
; APPLICANT: Prunkard, Donna E
; APPLICANT: Foster, Donald C
; TITLE OF INVENTION: Production of Fibrinogen in Transgenic
; TITLE OF INVENTION: Animals
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 4225 Roosevelt Way, N.E.
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/206,176
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31-648
; REFERENCE/DOCKET NUMBER: 93-15
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-547-8080 ext 322
; TELEFAX: 206-548-2329
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ovine beta-lactoglobulin
US-08-206-176-7

Query Match 35.5%; Score 28.4; DB 1; Length 10807;
Best Local Similarity 62.9%; Pred. No. 0.65;
Matches 44; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Patent No. 5741957
GENERAL INFORMATION:
APPLICANT: Deboer, Herman A.
APPLICANT: Strijker, Rein
APPLICANT: Heyneker, Herbert L.
APPLICANT: Platenburg, Gerald
APPLICANT: Lee, Sang He
APPLICANT: Pieper, Frank
APPLICANT: Krimpenfort, Paul J.A.
TITLE OF INVENTION: Production of Recombinant Polypeptides
by Bovine Species and Transgenic Methods
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Crew
STREET: One Market Plaza, Steuart Tower, Suite 2000
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,333
FILING DATE: 05-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/077,788
FILING DATE: 15-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/895,956
FILING DATE: 15-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/619,131
FILING DATE: 27-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/444,745
FILING DATE: 01-DEC-1989
ATTORNEY/AGENT INFORMATION:
NAME: Liebescheutz, Joe O.
REGISTRATION NUMBER: 37,505
REFERENCE/DOCKET NUMBER: 16994-003123
TELEPHONE: 415-543-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 824 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-461-333-29

Query Match 34.5%; Score 27.6; DB 1; Length 824;
Best Local Similarity 63.6%; Pred. No. 0.67;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
QY 13 cagccatgaggtgcttctgtactagctcttctgtgctctgcagtgaggagaccagtcaca 72
|||||
Db 754 CAGCCATGAAGTGCCTCTGCTTGCCTGGCCCTGCCTGTGCGGTCCAGGCCA 813
QY 73 acttgg 78
| | |
Db 814 TCATCG 819

RESULT 5
US-08-461-167-29
; Sequence 29, Application US/08464167

Patent No. 6013857
GENERAL INFORMATION:
APPLICANT: Deboer, Herman A.
APPLICANT: Strijker, Rein
APPLICANT: Heyneker, Herbert L.
APPLICANT: Platenburg, Gerald
APPLICANT: Lee, Sang He
APPLICANT: Pieper, Frank
APPLICANT: Krimpenfort, Paul J.A.
TITLE OF INVENTION: Production of Recombinant Polypeptides
by Bovine Species and Transgenic Methods
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Crew
STREET: One Market Plaza, Steuart Tower, Suite 2000
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,167
FILING DATE: 05-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/077,788
FILING DATE: 15-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/895,956
FILING DATE: 15-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/619,131
FILING DATE: 27-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/444,745
FILING DATE: 01-DEC-1989
ATTORNEY/AGENT INFORMATION:
NAME: Liebescheutz, Joe O.
REGISTRATION NUMBER: 37,505
REFERENCE/DOCKET NUMBER: 16994-003124
TELEPHONE: 415-543-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 824 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-464-167-29

Query Match 34.5%; Score 27.6; DB 3; Length 824;
Best Local Similarity 63.6%; Pred. No. 0.67;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
QY 13 cagccatgaggtgcttctgtactagctcttctgtgctctgcagtgaggagaccagtcaca 72
|||||
Db 754 CAGCCATGAAGTGCCTCTGCTTGCCTGGCCCTGCCTGTGCGGTCCAGGCCA 813
QY 73 acttgg 78
| | |
Db 814 TCATCG 819

RESULT 6
US-09-158-313-29
; Sequence 29, Application US/09158313

Patent No. 6066725

GENERAL INFORMATION:

APPLICANT: Deboer, Herman A.

APPLICANT: Strijker, Rein

APPLICANT: Heyneker, Herbert L.

APPLICANT: Platenburg, Gerald

APPLICANT: Lee, Sang He

APPLICANT: Pieper, Frank

APPLICANT: Krimpenfort, Paul J.A.

TITLE OF INVENTION: Production of Recombinant Polypeptides

BY: Bovine Species and Transgenic Methods

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew

STREET: One Market Plaza, Steuart Tower, Suite 2000

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94105

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/158,313

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/476,798

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US-07/895,956

FILING DATE: 15-JUN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/619,131

FILING DATE: 27-NOV-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/444,745

FILING DATE: 01-DEC-1989

ATTORNEY/AGENT INFORMATION:

NAME: Liebescheutz, Joe O.

REGISTRATION NUMBER: 37,505

REFERENCE/DOCKET NUMBER: 16994-003125

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-543-9600

TELEFAX: 415-543-5043

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 824 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-158-313-29

Query Match 34.5%; Score 27.6; DB 3; Length 824;
Best Local Similarity 63.6%; Pred. No. 0.67;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

[illegible]

RESULT 7
US-08-476-798-29
: Sequence 29, Application US/08476798

Patent NO. 6140552

GENERAL INFORMATION:

APPLICANT: Deboer, Herman A.

APPLICANT: Strijker, Rein

APPLICANT: Heyneker, Herbert L.

APPLICANT: Platenburg, Gerald

APPLICANT: Lee, Sang He

APPLICANT: Pieper, Frank

APPLICANT: Krimpenfort, Paul J.A.

TITLE OF INVENTION: Production of Recombinant Polypeptides

TITLE OF INVENTION: by Bovine Species and Transgenic Methods

NUMBER OF SEQUENCES: 36

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew

STREET: One Market plaza, Steuart Tower, Suite 2000

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94105

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/476,798

FILING DATE: 07-JUN-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/077,788

FILING DATE: 15-JUN-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/895,956

FILING DATE: 15-JUN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/619,131

FILING DATE: 27-NOV-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/444,745

FILING DATE: 01-DEC-1989

ATTORNEY/AGENT INFORMATION:

NAME: Liebeschultz, Joe O.

REGISTRATION NUMBER: 37,505

REFERENCE/DOCKET NUMBER: 16994-003125

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-543-9600

TELEFAX: 415-543-5043

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 824 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-476-798-29

Query Match 34.5%; Score 27.6; DB 4; Length 824;
Best Local Similarity 63.6%; Pred. No. 0.67;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Oy	13	cagccatgagggtcctgtactagctcttctgtggtctctcgatgggggacacagtcca	72
Db	754	CAGCCATGAAGTGCCTCCTCTGTCCTGGGCCCTCGCCTCGGTCCACGCCA	813
Oy	73	acttg	78
Db	814	TCATCG	819

RESULT 8
US-08-753-007A-3/c
; Sequence 3, Application US/08753007A


```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

RESULT 12
us-08-899-336-2/c
; Sequence 2, Application US/08899336
; Patent No. 5955649
; GENERAL INFORMATION:
; APPLICANT: HIROTA, NAOHIKO
; APPLICANT: KIHARA, MAKOTO
; APPLICANT: KURODA, HISAO
; APPLICANT: ITO, KAZUTOSHI
; TITLE OF INVENTION: GENE EXPRESSION REGULATORY DNA,
; TITLE OF INVENTION: GENE EXPRESSION CASSETTE, EXPRESSION VECTOR AND TRANSGENIC
; TITLE OF INVENTION: PLANT
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 SOUTH JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,336
; FILING DATE: 23-JUL-1997
; CLASSIFICATION: 800
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP HEI 8-193433
; FILING DATE: 23-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.


```

; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 274 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..274
; US-08-444-818-63

```

```

Query Match      29.5%; Score 23.6; DB 4; Length 274;
Best Local Similarity 61.3%; Pred. No. 12;
Matches 38; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 17 catgagggtgttactagctctgtggtctctcagtgaggaccagtcacatt 76
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 248 CAAGTGGTTCTATGGAGTAGGAGGCCCGGTAGATCTCGCAATCGAGGCGCTGTTCAGCT 189
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 77 99 78
   ||
Db 188 GG 187

```

Search completed: May 5, 2001, 22:04:58
Job time: 2397 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2001, 21:25:56 ; Search time 114.18 Seconds
(without alignments)
409.025 Million cell updates/sec

Title: US-09-426-776-11
Perfect score: 80
Sequence: 1 attaccatccaccagccatg.....gggaccagtcacaaattgggg 80

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401.*
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID56/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80	100.0	80	21	A28495 O. aureus vitellog
2	80	100.0	152	21	A28499 Vtgss-CAT fusion p
3	80	100.0	152	21	A28500 Vtgss-CAT fusion p
4	80	100.0	155	21	A28502 Vtgss-EGFP fusion
5	80	100.0	204	21	A28496 Vtgss-CrFCEs fusio
6	65.4	81.8	204	21	A28507 Vtgss-beta lactama
7	28.8	36.0	573	21	C79853 Human secreted pro
8	28.4	35.5	10807	16	T03855 Sheep beta-lactogl
9	28.4	35.5	10807	18	T79725 Ovine beta-lactogl
10	27.6	34.5	222	18	T60633 Human beta-lactogl
11	27.6	34.5	824	21	Z87633 Sheep beta-lactogl

c	12	27.6	34.5	824	22	C68328	Sheep beta lactogl
	13	27.4	34.2	879	21	F16035	Human prostate can
	14	26.8	33.5	4094	21	F51205	Human hypoxia regu
	15	26.8	33.5	4248	21	A28450	Human Seladin-1 CD
	16	26.6	33.2	2712	21	C46789	Arabidopsis thalia
	17	25.4	31.8	29	21	A28506	Primer BspSSfor fo
c	18	25.4	31.8	1607	19	V17813	Mus musculus don-1
c	19	25.4	31.8	2467	19	V17812	Mus musculus don-1
c	20	25	31.2	987	18	T89157	Oerskovia xanthine
c	21	25	31.2	1177	18	T89155	Oerskovia xanthine
c	22	25	31.2	1516	18	T89156	Oerskovia xanthine
	23	24.8	31.0	868	21	C46522	Zea mays DNA fragm
	24	24.8	31.0	1190	13	Q21979	Hybrid mini-vitell
	25	24.8	31.0	1247	13	Q21978	Hybrid mini-vitell
	26	24.8	31.0	1260	13	Q21976	Hybrid mini-vitell
	27	24.8	31.0	1290	13	Q21977	Hybrid mini-vitell
	28	24.8	31.0	1290	13	Q21981	Hybrid mini-vitell
	29	24.4	30.5	3394	21	D00315	Human Ras signall
	30	24.4	30.5	4109	21	D00316	Human CAMP-GRFI al
	31	24.2	30.2	1520	22	F32777	Human secreted pro
c	32	24.2	30.2	2296	19	V20662	Barley D-hordein g
c	33	23.8	29.8	1182	22	F4982	Murine INTERCEPT 2
c	34	23.8	29.8	1182	22	F45050	Murine secreted pr
c	35	23.8	29.8	1182	22	F45051	Murine secreted pr
c	36	23.8	29.8	1182	22	F45052	Murine secreted pr
c	37	23.8	29.8	1182	22	F45053	Murine secreted pr
c	38	23.8	29.8	1846	22	F4981	Murine INTERCEPT 2
c	39	23.8	29.8	1846	22	F45018	Murine secreted pr
c	40	23.8	29.8	1846	22	F45019	Murine secreted pr
c	41	23.8	29.8	1846	22	F45020	Murine secreted pr
c	42	23.8	29.8	1846	22	F45021	Murine secreted pr
c	43	23.8	29.8	2018	20	V68216	HOXB4 cDNA sequenc
c	44	23.8	29.8	2030	21	Z87687	Rat heart malonyl
c	45	23.8	29.8	2030	21	Z87688	Rat liver malonyl

ALIGNMENTS

RESULT 1
A28495 ID A28495 standard; DNA; 80 BP.
XX AC A28495;
XX DT 29-AUG-2000 (first entry)
XX DE O. aureus vitellogenin secretory signal sequence coding sequence.
XX VTgss; vitellogenin; secretory signal sequence; gene expression;
KW oestrogen receptor binding protein; systemic circulation; ss.
XX OS Oreochromis aureus.
XX FH Key Location/Qualifiers
FT sig_peptide 18..80
FT /*tag= a
XX PN WO200026366-A1.
XX PD 11-MAY-2000.
XX PF 29-OCT-1999; 99WO-SG00108.
XX PR 30-OCT-1998; 98US-0106426.
XX PR 26-OCT-1999; 99US-0426776.
XX XX (UYSI-) UNIV SINGAPORE NAT.
PA (LAMT/) LAM T J.
XX Ding JL, Tan NS, Ho B;
XX WPI; 2000-365615/31.
DR

DR P-PSDB; Y92778.

XX Isolated nucleic acid for assaying for heterologous gene expression,

PT detecting presence of compound that binds to estrogen receptor or

PT producing desired protein from host cell comprises nucleotide sequence

PT encoding secretory signal sequence

XX Claim 2; Page 36; 73pp; English.

XX The present sequence encodes a piscine, Oreochromis aureus, vitellogenin

CC secretory sequence (Vtgs). This and variants that comprise conservative

CC replacements that retain the biological activities of directing secretion

CC of a fusion protein from a cell and cleavage of the secretory signal

CC sequence from the fusion protein, are new. DNA encoding the Vtgs can be

CC fused to either a reporter protein or a lipopolysaccharide-binding

CC protein coding sequence. The isolated nucleic acid is useful for assaying

CC for heterologous gene expression, detecting the presence of a compound

CC that binds to an oestrogen receptor in a sample or producing a desired

CC protein from a host cell. It can also be used in a method for obtaining

CC systemic circulation of a desired protein in a transgenic or chimeric

XX host organism.

SQ Sequence 80 BP; 14 A; 22 C; 24 G; 20 T; 0 other;

Query Match 100.0%; Score 80; DB 21; Length 80;

Best Local Similarity 100.0%; Pred. No. 5.7e-18;

Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attcacatccaccagccatgaggggtgtgtactagcttctgtgtgctctcgcagtgg 60

DB 1 attcacatccaccagccatgaggggtgtgtactagcttctgtgtgctctcgcagtgg 60

OY 61 gggaccagtcacacttg99g 80

DB 61 gggaccagtcacacttg99g 80

RESULT 2

ID A28499 standard; DNA; 152 BP.

XX A28499;

XX 29-AUG-2000 (first entry)

XX Vtgs-CAT fusion protein (partial) coding sequence.

XX Vtgs; vitellogenin; secretory signal sequence; gene expression;

KW oestrogen receptor binding protein; systemic circulation; CAT; ss.

XX Chimeric - Oreochromis aureus.

OS Chimeric - Synthetic.

XX Key Location/Qualifiers

FT CDS 51..152

FT /*tag= a

FT /transl_except= (pos:123..125, aa:Gln)

FT /transl_except= (pos:132..134, aa:Val)

FT /transl_except= (pos:141..143, aa:Trp)

FT sig_peptide 51..113

FT /*tag= b

FT mat_peptide 117..152

FT /*tag= c

FT /partial

FT /note= "CAT N-terminal"

XX WO200026366-A1.

XX 11-MAY-2000.

XX 29-OCT-1999; 99WO-SG00108.

PR 30-OCT-1998; 98US-0106425.

PR 26-OCT-1999; 99US-0426776.

PA (UYSI-) UNIV SINGAPORE NAT.

PA (LAWT/) LAM T J.

XX Ding JL, Tan NS, Ho B;

XX WPI; 2000-365615/31.

DR P-PSDB; Y92780.

XX Isolated nucleic acid for assaying for heterologous gene expression,

PT detecting presence of compound that binds to estrogen receptor or

PT producing desired protein from host cell comprises nucleotide sequence

PT encoding secretory signal sequence

XX Example 2; Fig 5B; 73pp; English.

XX A reporter CAT system that uses the piscine, Oreochromis aureus,

CC vitellogenin secretory sequence (Vtgs), designated pBSVtgcAT was

CC constructed. A28499 and A28722 comprise the 5'- and 3'-ends of the

CC construct insert. Vtgs and variants that comprise conservative

CC replacements that retain the biological activities of directing secretion

CC of a fusion protein from a cell and cleavage of the secretory signal

CC sequence from the fusion protein, are new. DNA encoding the Vtgs can be

CC fused to either a reporter protein or a lipopolysaccharide-binding

CC protein coding sequence. The isolated nucleic acid is useful for assaying

CC for heterologous gene expression, detecting the presence of a compound

CC that binds to an oestrogen receptor in a sample or producing a desired

CC protein from a host cell. It can also be used in a method for obtaining

CC systemic circulation of a desired protein in a transgenic or chimeric

XX host organism.

SQ Sequence 152 BP; 38 A; 38 C; 39 G; 37 T; 0 other;

Query Match 100.0%; Score 80; DB 21; Length 152;

Best Local Similarity 100.0%; Pred. No. 6.6e-18;

Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attcacatccaccagccatgaggggtgtgtactagcttctgtgtgctctcgcagtgg 60

DB 34 attcacatccaccagccatgaggggtgtgtactagcttctgtgtgctctcgcagtgg 93

OY 61 gggaccagtcacacttg99g 80

DB 94 gggaccagtcacacttg99g 113

RESULT 3

ID A28500 standard; DNA; 152 BP.

XX A28500;

XX 29-AUG-2000 (first entry)

XX Vtgs-CAT fusion protein (partial) coding sequence.

XX Vtgs; vitellogenin; secretory signal sequence; gene expression;

KW oestrogen receptor binding protein; systemic circulation; CAT; ss.

XX Chimeric - Oreochromis aureus.

OS Chimeric - Synthetic.

XX Key Location/Qualifiers

FT CDS 51..152

FT /*tag= a

FT /transl_except= (pos:123..125, aa:Gln)

FT /transl_except= (pos:132..134, aa:Val)

FT /transl_except= (pos:141..143, aa:Trp)

FT sig_peptide 51..113

FT /*tag= b

```

FT mat_peptide 117..152
FT /*tag= c
FT /partial
FT /note= "CAT N-terminal"
XX
XX WO200026366-A1.
XX
XX 11-MAY-2000.
XX
XX 29-OCT-1999; 99WO-SG00108.
XX
XX 30-OCT-1998; 98US-0106426.
XX
XX 26-OCT-1999; 99US-0426776.
XX
XX (UYSI-) UNIV SINGAPORE NAT.
XX (LAMT/) LAM T J.
XX
XX Ding JL, Tan NS, Ho B;
XX
XX WPI; 2000-365615/31.
XX P-PSDB; Y92780.
XX
XX Isolated nucleic acid for assaying for heterologous gene expression,
XX detecting presence of compound that binds to estrogen receptor or
XX producing desired protein from host cell comprises nucleotide sequence
XX encoding secretory signal sequence
XX
XX Example 2; Fig 5D; 73pp; English.
XX
XX A reporter CAT system that uses the piscine, Oreochromis aureus,
XX vitellogenin secretory sequence (Vtgss), designated pep-vtgCAR was
XX constructed. A28500 and A28723 comprise the 5'- and 3'-ends of the
XX construct insert. Vtgss and variants that comprise conservative
XX replacements that retain the biological activities of directing secretion
XX of a fusion protein from a cell and cleavage of the secretory signal
XX sequence from the fusion protein, are new. DNA encoding the Vtgss can be
XX fused to either a reporter protein or a lipopolysaccharide-binding
XX protein coding sequence. The isolated nucleic acid is useful for assaying
XX for heterologous gene expression, detecting the presence of a compound
XX that binds to an estrogen receptor in a sample or producing a desired
XX protein from a host cell. It can also be used in a method for obtaining
XX systemic circulation of a desired protein in a transgenic or chimeric
XX host organism.
XX
XX Sequence 152 BP; 39 A; 38 C; 39 G; 36 T; 0 other;

Query Match 100.0%; Score 80; DB 21; Length 152;
Best Local Similarity 100.0%; Pred. No. 6.6e-18;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attcacatccaccagccatgaggggtgtgttactagctctgtgtggtctgcagtg 60
   |||||||
Db 34 attcacatccaccagccatgaggggtgtgttactagctctgtgtggtctgcagtg 93
   |||||||
QY 61 gggaccaggtccaaacttggg 80
   |||||||
Db 94 gggaccaggtccaaacttggg 113
   |||||||

RESULT 4
A28502
ID A28502 standard; DNA; 155 BP.
XX
AC A28502;
XX
XX 29-AUG-2000 (first entry)
XX
XX Vtgss-EGFP fusion protein (partial) coding sequence.
XX
XX Vtgss; vitellogenin; secretory signal sequence; gene expression;
XX oestrogen receptor binding protein; systemic circulation; EGFP; ss.
XX

```

```

OS Chimeric - Oreochromis aureus.
OS Chimeric - Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 42..155 /*tag= a
XX FT 42..104 /*tag= b
XX FT sig_peptide /note= "O. aureus Vtgss"
XX FT mat_peptide 123..155 /*tag= c
XX FT /*note= "EGFP"
XX
XX WO200026366-A1.
XX
XX 11-MAY-2000.
XX
XX 29-OCT-1999; 99WO-SG00108.
XX
XX 30-OCT-1998; 98US-0106426.
XX
XX 26-OCT-1999; 99US-0426776.
XX
XX (UYSI-) UNIV SINGAPORE NAT.
XX (LAMT/) LAM T J.
XX
XX Ding JL, Tan NS, Ho B;
XX
XX WPI; 2000-365615/31.
XX P-PSDB; Y92782.
XX
XX Isolated nucleic acid for assaying for heterologous gene expression,
XX detecting presence of compound that binds to estrogen receptor or
XX producing desired protein from host cell comprises nucleotide sequence
XX encoding secretory signal sequence
XX
XX Example 3; Fig 8B; 73pp; English.
XX
XX A reporter GFP system that uses the piscine, Oreochromis aureus,
XX vitellogenin secretory sequence (Vtgss), designated pVtgEGFP was
XX constructed. Vtgss and variants that comprise conservative
XX replacements that retain the biological activities of directing secretion
XX of a fusion protein from a cell and cleavage of the secretory signal
XX sequence from the fusion protein, are new. DNA encoding the Vtgss can be
XX fused to either a reporter protein or a lipopolysaccharide-binding
XX protein coding sequence. The isolated nucleic acid is useful for assaying
XX for heterologous gene expression, detecting the presence of a compound
XX that binds to an oestrogen receptor in a sample or producing a desired
XX protein from a host cell. It can also be used in a method for obtaining
XX systemic circulation of a desired protein in a transgenic or chimeric
XX host organism.
XX
XX Sequence 155 BP; 30 A; 45 C; 49 G; 31 T; 0 other;

Query Match 100.0%; Score 80; DB 21; Length 155;
Best Local Similarity 100.0%; Pred. No. 6.6e-18;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attcacatccaccagccatgaggggtgtgttactagctctgtgtggtctgcagtg 60
   |||||||
Db 25 attcacatccaccagccatgaggggtgtgttactagctctgtgtggtctgcagtg 84
   |||||||
QY 61 gggaccaggtccaaacttggg 80
   |||||||
Db 85 gggaccaggtccaaacttggg 104
   |||||||

RESULT 5
A28496
ID A28496 standard; DNA; 204 BP.
XX
XX AC A28496;
XX

```

DT 29-AUG-2000 (first entry)
 XX Vtgss-CrFCS fusion protein coding sequence.
 DE Vtgss; vitellogenin; secretory signal sequence; gene expression;
 KW oestrogen receptor binding protein; systemic circulation; factor C; ss.
 KW Chimeric - Oreochromis aureus.
 OS Chimeric - Carcinocorpius rotundicauda.
 XX Key Location/Qualifiers
 FT CDS 52..204
 FT /*tag= a
 FT /transl_except= (pos:199..201, aa:Tyr)
 FT /transl_except= (pos:202..204, aa:Phe)
 FT /*partial
 FT sig_peptide 52..114
 FT /*tag= b
 FT /*note= "O. aureus vitellogenin secretory signal sequence"
 FT mat_peptide 115..204
 FT /*tag= c
 FT /*note= "C. rotundicauda Factor C LPS-binding domain"
 PN WO200026366-A1.
 XX 11-MAY-2000.
 PD 29-OCT-1999; 99WO-SG00108.
 XX 30-OCT-1998; 98US-0106426.
 PR 26-OCT-1999; 99US-0426776.
 XX (UYSI-) UNIV SINGAPORE NAT.
 PA (LAMT/) LAM T J.
 PI Ding JL, Tan NS, Ho B;
 XX WPI; 2000-365615/31.
 DR P-PSDB; Y92779.
 XX Isolated nucleic acid for assaying for heterologous gene expression,
 PT detecting presence of compound that binds to estrogen receptor or
 PT producing desired protein from host cell comprises nucleotide sequence
 PT encoding secretory signal sequence
 XX Example 1; Fig 2A; 73pp; English.
 XX The piscine, Oreochromis aureus, vitellogenin secretory sequence (Vtgss)
 CC was fused upstream to the EcoRI-SalI CDNA fragment encoding the
 CC lipopolysaccharide-binding domain of Carcinocorpius rotundicauda Factor
 CC C CrFCS for expression and secretion of recombinant ES protein from
 CC Drosophila cells. Vtgss and variants that comprise conservative
 CC replacements that retain the biological activities of directing secretion
 CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Vtgss can be
 CC fused to either a reporter protein or a lipopolysaccharide-binding
 CC protein coding sequence. The isolated nucleic acid is useful for assaying
 CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an oestrogen receptor in a sample or producing a desired
 CC protein from a host cell. It can also be used in a method for obtaining
 CC systemic circulation of a desired protein in a transgenic or chimeric
 CC host organism.
 XX Sequence 204 BP; 48 A; 45 C; 58 G; 53 T; 0 other;
 SQ

Query Match 100.0%; Score 80; DB 21; Length 204;
 Best Local Similarity 100.0%; Pred. No. 7e-18;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 attcacatccaccagccatgaggggtgtgtactagcttctgtgctctcgcagtg 60
 DB 35 attcacatccaccagccatgaggggtgtgtactagcttctgtgctctcgcagtg 94

QY 61 gggaccagtcacacttgagg 80
 DB 95 gggaccagtcacacttgagg 114
 RESULT 6
 A28507
 ID A28507 standard; DNA; 204 BP.
 XX AC A28507;
 XX 29-AUG-2000 (first entry)
 XX Vtgss-beta-lactamase fusion protein (partial) coding sequence.
 DE Vtgss; vitellogenin; secretory signal sequence; gene expression;
 KW oestrogen receptor binding protein; systemic circulation;
 KW beta-lactamase; ss.
 XX Chimeric - Oreochromis aureus.
 OS Chimeric - Synthetic.
 XX Key Location/Qualifiers
 FT CDS 52..204
 FT /*tag= a
 FT /*partial
 FT sig_peptide 52..114
 FT /*tag= b
 FT /*note= "Vtgss from O. aureus"
 FT mat_peptide 115..204
 FT /*tag= c
 FT /*note= "beta-lactamase mature protein"
 XX WO200026366-A1.
 XX 11-MAY-2000.
 PD 29-OCT-1999; 99WO-SG00108.
 XX 30-OCT-1998; 98US-0106426.
 PR 26-OCT-1999; 99US-0426776.
 XX (UYSI-) UNIV SINGAPORE NAT.
 PA (LAMT/) LAM T J.
 PI Ding JL, Tan NS, Ho B;
 XX WPI; 2000-365615/31.
 DR P-PSDB; Y92783.
 XX Isolated nucleic acid for assaying for heterologous gene expression,
 PT detecting presence of compound that binds to estrogen receptor or
 PT producing desired protein from host cell comprises nucleotide sequence
 PT encoding secretory signal sequence
 XX Example 6; Fig 14A; 73pp; English.
 XX A reporter beta-lactamase system that uses the Oreochromis aureus,
 CC vitellogenin secretory sequence (Vtgss), designated pADVtgblactKana was
 CC constructed. Vtgss and variants that comprise conservative
 CC replacements that retain the biological activities of directing secretion
 CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Vtgss can be
 CC fused to either a reporter protein or a lipopolysaccharide-binding
 CC protein coding sequence. The isolated nucleic acid is useful for assaying
 CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an oestrogen receptor in a sample or producing a desired
 CC protein from a host cell. It can also be used in a method for obtaining
 CC systemic circulation of a desired protein in a transgenic or chimeric
 CC host organism.
 XX Sequence 204 BP; 49 A; 42 C; 60 G; 53 T; 0 other;
 SQ

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2001, 15:51:41 ; Search time 8.3 Seconds
(without alignments)
86.671 Million cell updates/sec

Title: US-09-426-776-10

Perfect score: 96

Sequence: 1 MRVLVLAVALAVGDSNLG 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	57.3	674	1 GSPD_VIBCH	P45779 vibrio chol
2	53	55.2	678	1 GSPD_AERHY	P31780 aeromonas h
3	53	55.2	678	1 GSPD_AERSA	P45778 aeromonas s
4	53	55.2	1687	1 VIT2_FUNHE	Q98893 fundulus he
5	45	46.9	945	1 ATIK_HUMAN	O60423 homo sapien
6	44	45.8	629	1 ASM_HUMAN	P17405 homo sapien
7	43	44.8	222	1 CASB_CAPHI	P33048 capra hircu
8	43	44.8	262	1 APAL_BRARE	O42363 brachydanio
9	43	44.8	683	1 BGH3_HUMAN	Q15582 homo sapien
10	43	44.8	683	1 BGH3_RABIT	Q95215 oryctolagus
11	43	44.8	1148	1 AT1B_MOUSE	P98200 mus musculu
12	43	44.8	1251	1 AT1C_HUMAN	O43520 homo sapien
13	42.5	44.3	80	1 YJBE_ECOLI	P32686 escherichia
14	42	43.8	168	1 PBP_MANSE	P18959 manduca sex
15	42	43.8	258	1 APAL_SALSA	P27007 salmo salar
16	42	43.8	378	1 CYOB_HAEIN	P45020 haemophilus
17	42	43.8	550	1 YCHM_ECOLI	P40877 escherichia
18	42	43.8	654	1 GSPD_ECOLI	P45758 escherichia
19	41	42.7	180	1 LACB_CAPHI	P02756 capra hircu
20	41	42.7	180	1 LACB_SHEEP	P02757 ovis aries
21	41	42.7	222	1 CASB_SHEEP	P11839 ovis aries
22	41	42.7	224	1 CASE_BOVIN	P02666 bos taurus
23	41	42.7	224	1 CASE_BUBBU	Q9ts10 bubalus bub
24	41	42.7	226	1 CASE_HUMAN	P05814 homo sapien
25	41	42.7	228	1 CASB_RABIT	P09116 oryctolagus
26	41	42.7	232	1 CASB_CAMDR	Q9tvd0 camelus dro
27	41	42.7	312	1 HO2_RABIT	P43242 oryctolagus
28	41	42.7	387	1 PAB_PEPMA	Q51911 peptococci
29	41	42.7	397	1 MML6_MYCTU	Q10773 mycobacteri
30	41	42.7	755	1 COMP_RAT	P35444 rattus norv
31	41	42.7	958	1 MML1_MYCTU	P95211 mycobacteri
32	41	42.7	1122	1 ADPL_MYCGA	O49379 mycoplasma
33	41	42.7	1163	1 AT5C_HUMAN	O60312 homo sapien

34 41 42.7 1508 1 AT5A_MOUSE
35 40.5 42.2 81 1 ATPL_SYNP6
36 40 41.7 160 1 GLB2_CHITH
37 40 41.7 166 1 PAL_PSEPU
38 40 41.7 185 1 UPK2_BOVIN
39 40 41.7 258 1 IBP4_HUMAN
40 40 41.7 266 1 PTNC_ECOLI
41 40 41.7 322 1 SERE_ECOLI
42 40 41.7 362 1 ALEU_HORVU
43 40 41.7 499 1 PITA_ECOLI
44 40 41.7 499 1 PITA_ECOLI
45 40 41.7 516 1 GUX1_NEUCR

ALIGNMENTS

RESULT 1
GSPD_VIBCH STANDARD; PRT; 674 AA.
AC P45779;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GENERAL SECRETION PATHWAY PROTEIN D PRECURSOR (CHOLERA TOXIN SECRETION
DE PROTEIN EPSD).
GN EPSD OR VC2733.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR TRH7000;
RA Overbye L.J.;
RT "Organization of the general secretion pathway genes in Vibrio
RT cholerae".
RL Thesis (1994), Michigan State University / East Lansing, U.S.A.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardon D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae".
RL Nature 406:477-483(2000).
CC -!- FUNCTION: REQUIRED FOR SECRETION OF CHOLERA TOXIN THROUGH THE
CC OUTER MEMBRANE.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE EXED/OUTD/FULD/XPSD FAMILY.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT (POSSIBLY NATURAL) IN POSITION 55.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L33796; AAA58785.1; -;
CC EMBL: AB004338; -; NOT_ANNOTATED_CDS.
CC TIGR: VC2733; -;
CC InterPro: IPR00016; -;
CC InterPro: IPR001775; -;
CC Pfam: PF00263; Bac_GSPproteins; 1.
CC PRINTS: PR00811; BCTERIALGSPD.

Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 RVLVLALAVAVLAVGDSNGLG 21
 :||| | : | ||| |||
 Db 353 QVLVEAIVEIADGDLNGLG 372

RESULT 3
 GSPD_AERSA STANDARD; PRT; 678 AA.
 ID AC P45778;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE GENERAL SECRETION PATHWAY PROTEIN D PRECURSOR.
 GN EXED.
 OS Aeromonas salmonicida.
 OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
 OC Aeromonas.
 OC NCBI_TaxID=645;
 RN [1].
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCMB 1102;
 RX MEDLINE=95309729; PubMed=7789814;
 RA Karlyshev A.V., Macintyre S.;
 RT "Cloning and study of the genetic organization of the exo gene
 cluster of Aeromonas salmonicida.";
 RL Gene 158777-82(1995).
 CC -1- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
 CC EXPORT OF PROTEINS.
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE EXED/OUTD/EULD/XPSD FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X80505; CAA56668.1; -
 CC InterPro; IPR000016; -
 CC DR InterPro; IPR001775; -
 CC DR Pfam; PF00263; Bac_GSPproteins; 1.
 CC DR PRINTS; PR00811; BCTERIALGSPD.
 CC DR PROSITE; PS00875; T2SP_D; 1.
 KW Transport; Outer membrane; Signal.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 678
 FT SEQUENCE 678 AA; 72768 MW; CB4921C9BAA8438E CRC64;
 SQ GENERAL SECRETION PATHWAY PROTEIN D.

Query Match 55.2%; Score 53; DB 1; Length 678;
 Best Local Similarity 60.0%; Pred. No. 4;
 Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 RVLVLALAVAVLAVGDSNGLG 21
 :||| | : | ||| |||
 Db 352 QVLVEAIVEIADGDLNGLG 371

RESULT 4
 VIT2_FUNHE STANDARD; PRT; 1687 AA.
 ID AC Q98893;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE VITELLOGENIN II PRECURSOR (VTG II) [CONTAINS: LIPOVITELLIN 1 (LVI);
 DE PHOSVITIN (PV); LIPOVITELLIN 2 (LV2); YP 69].
 OS Fundulus heteroclitus (Killifish) (Mummichog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
CC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
CC Cyprinodontiformes; Fundulidae; Fundulus.
OX NCBI_TaxID=8078;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-35.
RC TISSUE=Liver;
RA Lafleur G.J. Jr., Byrne B.M., Haux C., Greenberg R.M., Wallace R.A.;
RT "Liver-derived cDNAs: vitellogenins and vitelline envelope protein
precursors (choriogenins).";
RL Int. Symp. Reprod. Physiol. Fish 5:336-338(1995).
CC -1- FUNCTION: PRECURSOR OF THE EGG-YOLK PROTEINS THAT ARE
CC SOURCES OF NUTRIENTS DURING EARLY DEVELOPMENT OF OVIPAROUS
CC ORGANISMS.
CC -1- TISSUE SPECIFICITY: PRODUCED BY THE LIVER, SECRETED INTO THE BLOOD
CC AND THEN SEQUESTERED BY RECEPTOR MEDIATED ENDOCYTOSIS INTO GROWING
CC OOCYTES, WHERE IT IS GENERALLY CLEAVED, GIVING RISE TO THE
CC RESPECTIVE YOLK COMPONENTS LIPOVITELLINS AND PHOSVITIN.
CC -1- INDUCTION: SYNTHESIZED IN THE LIVER OF OVIPAROUS VERTEBRATES IN
CC RESPONSE TO STEROID (ESTROGEN) INDUCTION. STEROID INDUCED
CC EXPRESSION OF VTG II IS LOWER THAN THAT OF VTG I.
CC -1- PTM: PHOSVITIN, AN EGG YOLK STORAGE PROTEIN, IS ONE OF THE MOST
CC HIGHLY PHOSPHORYLATED (10%) PROTEINS IN NATURE (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U70826; AB17152.1; -
CC InterPro; IPR001747; -
CC DR InterPro; IPR001846; -
CC DR Pfam; PF01347; Vitellogenin_N; 1.
CC DR Pfam; PF00094; vwd; 1.
CC KW Glycoprotein; phosphorylation; Storage protein; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 1687 VITELLOGENIN II.
FT CHAIN 16 ? LIPOVITELLIN 1.
FT CHAIN ? ? PHOSVITIN.
FT CHAIN ? 1687 LIPOVITELLIN 2.
FT DOMAIN 1059 1062 POLY-GLU.
FT DOMAIN 1088 1169 SER-RICH.
FT CARBOHYD 941 941 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 945 945 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 954 954 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1004 1004 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1019 1019 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1083 1083 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1142 1142 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1179 1179 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1257 1257 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1292 1292 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1342 1342 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1361 1361 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1366 1366 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1390 1390 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1577 1577 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1655 1655 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1687 AA; 186005 MW; 4965BB9DBFB4928F CRC64;
Query Match 55.2%; Score 53; DB 1; Length 1687;
Best Local Similarity 75.0%; Pred. No. 8.2;
Matches 12; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 MRVLVLAVALAVGD 16
Db 1 MRVLVLAVALVAGN 16

RESULT 5
AT1K_HUMAN STANDARD; PRT; 945 AA.
AC 060423;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE POTENTIAL PHOSPHOLIPID-TRANSPORTING ATPASE IK (EC 3.6.1.-) (FRAGMENT).
GN POS37502.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schulz K., Gordon L., Ramirez M., Stilwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Krommiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP CONCEPTUAL TRANSLATION.
RA Axelsten K.B.;
RL Unpublished observations (FEB-2000).
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASE FAMILY (EI-E2
CC ATPASES). SUBFAMILY IV.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS AT SEVERAL POSITIONS FROM THAT
CC SHOWN DUE TO DIFFERENCES IN THE PREDICTION OF SPLICE SITES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AC004755; AAC17601.1; ALT_SEQ.
CC InterPro; IPR001757; -
CC DR InterPro; IPR002048; -
CC DR Pfam; PF00122; EI-E2_ATPase; 2.
CC DR PROSITE; PS00154; ATPASE_EI_E2; 1.
CC KW Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
CC Multigene family.
FT NON_TER 1 1 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1 148 POTENTIAL.
FT TRANSMEM 149 170 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 171 197 POTENTIAL.
FT TRANSMEM 198 219 POTENTIAL.
FT DOMAIN 220 767 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 768 788 POTENTIAL.
FT DOMAIN 789 800 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 801 820 POTENTIAL.
FT DOMAIN 821 850 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 851 872 POTENTIAL.
FT DOMAIN 873 884 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 885 907 POTENTIAL.
FT DOMAIN 908 913 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 914 934 POTENTIAL.
FT DOMAIN 930 933 POLY-THR.
FT DOMAIN 935 945 EXTRACELLULAR (POTENTIAL).
FT MOD_RES 262 262 PHOSPHORYLATION (BY SIMILARITY).
FT METAL 711 711 MAGNESIUM (BY SIMILARITY).
FT METAL 715 715 MAGNESIUM (BY SIMILARITY).
FT NON_TER 945 945
SQ SEQUENCE 945 AA; 107157 MW; 2A28483C2F58D702 CRC64;

Query Match 46.9%; Score 45; DB 1; Length 945;
 Best Local Similarity 41.7%; Pred. No. 63;
 Matches 10; Conservative 5; Mismatches 3; Indels 6; Gaps 1;

OY 2 RVLVLAL-----AVALVDCGNS 19

Db 691 KALIVALVKYHOVVTLAIGDGN 714

RESULT 6

ASM_HUMAN STANDARD; PRT; 629 AA.

AC PI7405; PI7406;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE SPHINGOMYELIN PHOSPHODIESTERASE PRECURSOR (EC 3.1.4.12) (ACID
 DE SPHINGOMYELINASE).
 GN SMPDI OR ASM.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]

SEQUENCE FROM N.A.
 RX MEDLINE-91217097; PubMed-1840600;
 RA Schuchman E.H., Suchi M., Takahashi T., Sandhoff K., Desnick R.J.;
 RT "Human acid sphingomyelinase. Isolation, nucleotide sequence and
 RT expression of the full-length and alternatively spliced cDNAs.";
 RL J. Biol. Chem. 266:8531-8539(1991).
 RN [2]

SEQUENCE FROM N.A.
 RX MEDLINE-93183402; PubMed-1292508;
 RA Newzella D., Stoffel W.;
 RT "Molecular cloning of the acid sphingomyelinase of the mouse and the
 RT organization and complete nucleotide sequence of the gene.";
 RL Biol. Chem. Hoppe-Seyler 373:1233-1238(1992).
 RN [3]

SEQUENCE OF 128-629 FROM N.A., AND PARTIAL SEQUENCE.
 RP TISSUE-Fibroblast;
 RX MEDLINE-90060003; PubMed-2555181;
 RA Quintern L.E., Schuchman E.H., Levran O., Suchi M., Ferlinz K.,
 RA Reinke H., Sandhoff K., Desnick R.J.;
 RT "Isolation of cDNA clones encoding human acid sphingomyelinase:
 RT occurrence of alternatively processed transcripts.";
 RL EMBO J. 8:2469-2473(1989).
 RN [4]

VARIANT NPD SER-577.
 RP Ferlinz K., Hurwitz R., Sandhoff K.;
 RA "Molecular basis of acid sphingomyelinase deficiency in a patient
 RT with Niemann-Pick disease type A.";
 RL Biochem. Biophys. Res. Commun. 179:1187-1191(1991).
 RN [5]

VARIANT NPD LEU-496.
 RX MEDLINE-91219449; PubMed-2023926;
 RA Levran O., Desnick R.J., Schuchman E.H.;
 RT "Niemann-Pick disease: a frequent missense mutation in the acid
 RT sphingomyelinase gene of Ashkenazi Jewish type A and B patients.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3748-3752(1991).
 RN [6]

VARIANT NPD ARG-608 DEL.
 RX MEDLINE-91358737; PubMed-1885770;
 RA Levran O., Desnick R.J., Schuchman E.H.;
 RT "Niemann-Pick type B disease. Identification of a single codon
 RT deletion in the acid sphingomyelinase gene and genotype/phenotype
 RT correlations in type A and B patients.";
 RL J. Clin. Invest. 88:806-810(1991).
 RN [7]

VARIANT NPD PRO-302.
 RX MEDLINE-93004773; PubMed-1391960;
 RA Levran O., Desnick R.J., Schuchman E.H.;

*Identification and expression of a common missense mutation (L302P)
 in the acid sphingomyelinase gene of Ashkenazi Jewish type A
 Niemann-Pick disease patients.";
 RL Blood 80:2081-2087(1992).
 RN [8]

VARIANT NPD ARG-436.
 RX MEDLINE-93244834; PubMed-1301192;
 RA Takahashi T., Desnick R.J., Takada G., Schuchman E.H.;
 RT "Identification of a missense mutation (S436R) in the acid
 RT sphingomyelinase gene from a Japanese patient with type B
 RT Niemann-Pick disease.";
 RL Hum. Mutat. 1:70-71(1992).
 RN [9]

VARIANTS NPD ARG-242; ILE-382 AND SER-383.
 RX MEDLINE-92316934; PubMed-1618760;
 RA Takahashi T., Suchi M., Desnick R.J., Takada G., Schuchman E.H.;
 RT "Identification and expression of five mutations in the human acid
 RT sphingomyelinase gene causing types A and B Niemann-Pick disease.
 RT Molecular evidence for genetic heterogeneity in the neuropathic and
 RT non-neuropathic forms.";
 RL J. Biol. Chem. 267:12552-12558(1992).
 RN [10]

VARIANT NPD GLY-391.
 RX MEDLINE-94328611; PubMed-8051942;
 RA Spierl W., Bart G., Vanier M.T., Christomanou H., Baldissera I.,
 RA Steichensdorf E., Paschke E.;
 RT "A family with visceral course of Niemann-Pick disease, macular halo
 RT syndrome and low sphingomyelin degradation rate.";
 RL J. Inher. Metab. Dis. 17:93-103(1994).
 RN [11]

VARIANT NPD THR-389.
 RX MEDLINE-96287387; PubMed-8680412;
 RA Schuchman E.H.;
 RT "Two new mutations in the acid sphingomyelinase gene causing type A
 RT Niemann-Pick disease: N389T and R441X.";
 RL Hum. Mutat. 6:352-354(1995).
 RN [12]

VARIANT NPD GLN-246.
 RX MEDLINE-96263741; PubMed-8664904;
 RA Ida H., Rennert O.M., Maekawa K., Eto Y.;
 RT "Identification of three novel mutations in the acid
 RT sphingomyelinase gene of Japanese patients with Niemann-Pick disease
 RT type A and B.";
 RL Hum. Mutat. 7:65-67(1996).
 RN [13]

FUNCTION: CONVERTS SPHINGOMYELIN TO CERAMIDE. ASM ALSO HAS
 PHOSPHOLIPASE C ACTIVITIES TOWARD 1,2-DIACYLGLYCEROLPHOSPHOCHOLINE
 AND 1,2-DIACYLGLYCEROLPHOSPHOGLYCEROL.
 CC -!- CATALYTIC ACTIVITY: SPHINGOMYELIN + H(2)O = N-ACYLSPHINGOSINE +
 CC CHOLINE PHOSPHATE.
 CC -!- SUBUNIT: MONOMER.
 CC -!- SUBCELLULAR LOCATION: LYSOSOMAL.
 CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; ASM-1 (SHOWN HERE)
 CC (FREQUENCY 90%), ASM-2 (FREQUENCY 10%) AND ASM-3 (FREQUENCY <1%);
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. ONLY ASM-1 ENCODES A
 CC CATALYTICALLY ACTIVE ENZYME.
 CC -!- DISEASE: DEFICIENCY IN ASM IS THE CAUSE OF THE NIEMANN-PICK
 CC DISEASE (NPD). SPHINGOMYELIN AND OTHER METABOLICALLY RELATED
 CC LIPIDS ACCUMULATE IN THE LYSOSOMES. PATIENTS MAY SHOW XANTHOMAS,
 CC PIGMENTATION, HEPATOSPLENOMEGALY, LYMPHADENOPATHY AND MENTAL
 CC RETARDATION. THERE ARE TWO MAIN TYPES: TYPE A, IS A RAPIDLY
 CC PROGRESSIVE AND FATAL NEURODEGENERATIVE DISEASE OF INFANCY.
 CC TYPE B, HAS LITTLE IF ANY NEUROLOGIC INVOLVEMENT AND MAY
 CC SURVIVE INTO ADULTHOOD. NPD OCCURS MORE FREQUENTLY AMONG
 CC INDIVIDUALS OF ASHKENAZI JEWISH ANCESTRY THAN IN THE GENERAL
 CC POPULATION.
 CC -!- MISCELLANEOUS: THERE ARE TWO TYPES OF SPHINGOMYELINASES: ASM
 CC (ACID), AND NSM (NEUTRAL).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

BA Babin P.J., Thisse C., Durliat M., Andre M., Akimenko M.-A.,
 RA Thisse B.;
 RT "Both apolipoprotein E and A-I genes are present in a nonmammalian
 RT vertebrate and are highly expressed during embryonic development.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:8622-8627(1997).
 CC -1- FUNCTION: APOA-1 PARTICIPATES IN THE REVERSE TRANSPORT OF
 CC CHOLESTEROL FROM TISSUES TO THE LIVER FOR EXCRETION BY PROMOTING
 CC CHOLESTEROL EFFLUX FROM TISSUES AND BY ACTING AS A COFACTOR FOR
 CC THE LECITHIN CHOLESTEROL ACYLTRANSFERASE (LCAT) (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- DEVELOPMENTAL STAGE: HIGHLY EXPRESSED IN THE YOLK SYNCYTIAL LAYER
 CC DURING EMBRYONIC (STARTING AT THE GASTRULA STAGE) AND EARLY LARVAL
 CC DEVELOPMENT, AN EXTRAEMBRYONIC STRUCTURE IMPLICATED IN EMBRYONIC
 CC AND LARVAL NUTRITION.
 CC -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y13653; CAA74004.1; -;
 DR ZFIN; ZDB-GENE-990415-14; apoa.
 DR InterPro; IPR000074; -;
 DR Pfam; PF01442; Apolipoprotein; 1.
 KW Plasma; Lipid transport; HDL; Cholesterol metabolism; Repeat; Signal.
 FT SIGNAL 1 18 POTENTIAL.
 FT PROPEP 19 23 POTENTIAL.
 FT CHAIN 24 262 APOLIPOPROTEIN A-I.
 FT DOMAIN 32 63 3 X APPROXIMATE TANDEM REPEATS.
 FT DOMAIN 64 262 10 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 64 85 1.
 FT REPEAT 87 107 2.
 FT REPEAT 108 118 3 (HALF-LENGTH).
 FT REPEAT 119 140 4.
 FT REPEAT 141 162 5.
 FT REPEAT 163 184 6.
 FT REPEAT 185 206 7.
 FT REPEAT 207 228 8.
 FT REPEAT 229 239 9 (HALF-LENGTH).
 FT REPEAT 240 262 10.
 SQ SEQUENCE 262 AA; 30256 MW; BB839A0A815365B9 CRC64;

Query Match 44.8%; Score 43; DB 1; Length 262;
 Best Local Similarity 45.0%; Pred. No. 42;
 Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
 QY 1 MRVLVLALAVAVGDSNLS 20
 I : I I I : I I I : I I
 Db 1 MKFVALATLLALGSQLN 20
 I : I I I : I I I : I I

RESULT 9
 BGH3_HUMAN STANDARD; PRT; 683 AA.
 ID BGH3_HUMAN
 AC Q15582; O43216; O43217; O43218; O43219; O14471; O14472; O14476;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE TRANSFORMING GROWTH FACTOR-BETA INDUCED PROTEIN IG-H3 PRECURSOR (BETA
 DE IG-H3) (KERATO-EPITHELIN) (RGD-CONTAINING COLLAGEN ASSOCIATED PROTEIN)
 DE (RGD-CAP).
 GN TGFBI OR BIGH3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=93000472; PubMed=1388724;
 RA Skonier J., Neubauer M., Madisen L., Bennett K., Plowman G.D.,
 RA Purchio A.F.;
 RT "cDNA cloning and sequence analysis of beta ig-h3, a novel gene
 RT induced in a human adenocarcinoma cell line after treatment with
 RT transforming growth factor-beta";
 RL DNA Cell Biol. 11:511-522(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kimmerly W., Bondoc M., Cheng J., Connolly K.S., Gunning K.M.,
 RA Davis C.A., Kadner K., Miguel T., Pitluck S., Pollard M., Rojeski H.,
 RA Subramanian S., Martin C.H.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP VARIANTS CORNEAL DYSTROPHIES CYS-124; HIS-124; GLN-555 AND TRP-555.
 RX MEDLINE=97207642; PubMed=9054935;
 RA Munier F.L., Korvatska E., Djemai A., le Paslier D., Zografos L.,
 RA Pescia G., Schorderet D.F.;
 RT "Kerato-epithelin mutations in four 5q31-linked corneal dystrophies";
 RL Nat. Genet. 15:247-251(1997).
 RN [4]
 RP VARIANTS CORNEAL DYSTROPHIES CYS-124; HIS-124; GLN-555 AND TRP-555.
 RX MEDLINE=98130534; PubMed=9463327;
 RA Korvatska E., Munier F.L., Djemai A., Wang M.X., Frueh B.,
 RA Chiou A.G.-Y., Uffer S., Ballesstrazzi E., Braunstein R.E.,
 RA Forster R.K., Culbertson W.W., Roman H., Zografos L., Schorderet D.F.;
 RT "Mutation hot spots in 5q31-linked corneal dystrophies";
 RL Am. J. Hum. Genet. 62:320-324(1998).
 RN [5]
 RP VARIANT CDL3A THR-501.
 RX MEDLINE=98163459; PubMed=9497262;
 RA Yamamoto S., Okada M., Tsujikawa M., Shimomura Y., Nishida K.,
 RA Inoue Y., Watanabe H., Maeda N., Kurahashi H., Kinoshita S.,
 RA Nakamura Y., Tano Y.;
 RT "A kerato-epithelin (beta-ig-h3) mutation in lattice corneal dystrophy
 RT type IIIA";
 RL Am. J. Hum. Genet. 62:719-722(1998).
 RN [6]
 RP VARIANT CDRB LEU-124.
 RX MEDLINE=98451378; PubMed=9780098;
 RA Okada M., Yamamoto S., Tsujikawa M., Watanabe H., Inoue Y., Maeda N.,
 RA Shimomura Y., Nishida K., Quantock A.J., Kinoshita S., Tano Y.;
 RT "Two distinct kerato-epithelin mutations in Reis-Bucklers corneal
 RT dystrophy";
 RL Am. J. Ophthalmol. 126:535-542(1998).
 RN [7]
 RP VARIANT CDLI ARG-527.
 RX MEDLINE=99013426; PubMed=9799082;
 RA Fujiki K., Hotta Y., Nakayasu K., Yokoyama T., Takano T.,
 RA Yamaguchi T., Kanai A.;
 RT "A new L527R mutation of the betaIGH3 gene in patients with lattice
 RT corneal dystrophy with deep stromal opacities";
 RL Hum. Genet. 103:286-289(1998).
 RN [8]
 RP VARIANT CDRB PHE-540 DEL.
 RA Rozzo C., Fossarello M., Galleri G., Serru A., Orzalessi N.,
 RA Serra A., Pirastu M.;
 RT "A common beta ig-h3 gene mutation (delta F540) in a large cohort of
 RT Sardinian Reis-Bucklers' corneal dystrophy patients";
 RL Hum. Mutat. 12:215-216(1998).
 CC -1- FUNCTION: BINDS TO TYPE I, II, AND IV COLLAGENS. THIS ADHESION
 CC PROTEIN MAY PLAY AN IMPORTANT ROLE IN CELL-COLLAGEN INTERACTIONS.
 CC IN CARTILAGE, MAY BE INVOLVED IN ENDOCHONDRAL BONE FORMATION.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR. MAY BE ASSOCIATED BOTH WITH
 CC MICROFIBRILS AND WITH THE CELL SURFACE.
 CC -1- INDUCTION: BY TGF-BETA.
 CC -1- DISEASE: DEFECTS IN TGFBI ARE THE CAUSE OF AT LEAST FIVE AUTOSOMAL
 CC DOMINANT TYPES OF CORNEAL DYSTROPHIES: GRANULAR DYSTROPHY GROENOUW
 CC TYPE I (CDGGL), REIS-BUECKERS CORNEAL DYSTROPHY (CDRB), LATTICE
 CC CORNEAL DYSTROPHY TYPES I AND IIIA (CDLI AND CDL3A) AND AVELLINO
 CC CORNEAL DYSTROPHY (ACD). CLINICALLY THEY SHOW PROGRESSIVE
 CC OPACIFICATION OF THE CORNEA LEADING TO SEVERE VISUAL HANDICAP.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M77349; AAA61163.1; --
DR EMBL; AC004503; AAC08449.1; --
DR EMBL; AC005219; AAC24944.1; --
DR EMBL; AF035626; AAB88695.1; --
DR EMBL; AF035627; AAB88698.1; --
DR EMBL; AF035628; AAB88696.1; --
DR EMBL; AF035629; AAB88697.1; --
DR MIM; 601692; --
DR MIM; 121900; --
DR MIM; 122200; --
DR MIM; 122200; --
KW Extracellular matrix; Signal; Repeat; Cell adhesion; Disease mutation.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 683 TRANSFORMING GROWTH FACTOR-BETA INDUCED
FT CHAIN 16 683 PROTEIN IG-H3.
FT DOMAIN 103 236 BIGH3 1.
FT DOMAIN 240 371 BIGH3 2.
FT DOMAIN 375 498 BIGH3 3.
FT DOMAIN 502 632 BIGH3 4.
FT SITE 642 644 CELL ATTACHMENT SITE (POTENTIAL).
FT VARIANT 124 124 R -> C (IN CDL1).
FT VARIANT 124 124 R -> H (IN ACD).
FT VARIANT 124 124 /FTID=VAR_005076.
FT VARIANT 124 124 /FTID=VAR_005077.
FT VARIANT 124 124 R -> L (IN CDRB).
FT VARIANT 501 501 /FTID=VAR_005078.
FT VARIANT 501 501 P -> T (IN CDL3A).
FT VARIANT 527 527 /FTID=VAR_005079.
FT VARIANT 527 527 L -> R (IN CDL1, LATE-ONSET).
FT VARIANT 540 540 /FTID=VAR_005080.
FT VARIANT 555 555 MISSING (IN CDRB).
FT VARIANT 555 555 /FTID=VAR_005081.
FT VARIANT 555 555 R -> Q (IN CDRB).
FT VARIANT 555 555 /FTID=VAR_005082.
FT VARIANT 555 555 R -> W (IN CDG1).
FT SEQUENCE 683 AA; 74680 MW; 40FDC8A71EBB3DD00 CRC64;

Query Match 44.8%; Score 43; DB 1; Length 683;
Best Local Similarity 55.6%; Pred. NO. 90;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 VLVLAVALAVGDSNLS 20
DB 5 VRLALALALGPAATL 22

RESULT 10

EGH3_RABIT
ID BGH3_RABIT STANDARD; PRT; 683 AA.
AC Q95215;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSFORMING GROWTH FACTOR-BETA INDUCED PROTEIN IG-H3 PRECURSOR (BETA
DE IG-H3) (KERATO-EPITHELIN) (RGD-CONTAINING-COLLAGEN ASSOCIATED PROTEIN)
DE (RGD-CAP).
GN TGFBI.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NEW ZEALAND WHITE; TISSUE-Cornea;

RX MEDLINE-97267655; PubMed-9112985;
RA Rave I.M., Zhan Q., Burrows R., Bennett K., Cintron C.;
RT "Beta-ig. Molecular cloning and in situ hybridization in corneal
RL tissues";
RL Invest. Ophthalmol. Vis. Sci. 38:893-900(1997).
CC -!- FUNCTION: BINDS TO TYPE I, II, AND IV COLLAGENS. THIS ADHESION
CC PROTEIN MAY PLAY AN IMPORTANT ROLE IN CELL-COLLAGEN INTERACTIONS.
CC IN CARTILAGE, MAY BE INVOLVED IN ENDOCHONDRAL BONE FORMATION. MAY
CC PLAY AN IMPORTANT ROLE IN THE MORPHOGENESIS OF CORNEAL TISSUES.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR. MAY BE ASSOCIATED BOTH WITH
CC MICROFIBRILS AND WITH THE CELL SURFACE (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: LOCATED PRIMARILY IN THE EPITHELIUM OF NORMAL
CC ADULT CORNEA, IN FETAL STROMAL CELLS, AND BOTH ENDOTHELIUM- AND
CC STROMA-DERIVED CELLS IN HEALING CORNEAL WOUNDS. NOT EXPRESSED IN
CC NORMAL ADULT ENDOTHELIUM AND STROMA.
CC -!- INDUCTION: BY TGF-BETA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; U66205; AAB07015.1; ALT_FRAME.
KW Extracellular matrix; Signal; Repeat; Cell adhesion.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 683 TRANSFORMING GROWTH FACTOR-BETA INDUCED
FT CHAIN 24 683 PROTEIN IG-H3.
FT DOMAIN 103 236 BIGH3 1.
FT DOMAIN 240 371 BIGH3 2.
FT DOMAIN 375 498 BIGH3 3.
FT DOMAIN 502 632 BIGH3 4.
FT SITE 641 643 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 683 AA; 74684 MW; 4548520497548CD6 CRC64;

Query Match 44.8%; Score 43; DB 1; Length 683;
Best Local Similarity 55.6%; Pred. NO. 90;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 VLVLAVALAVGDSNLS 20
DB 5 VRLALALALGPAATL 22

RESULT 11

AT1B_MOUSE
ID AT1B_MOUSE STANDARD; PRT; 1148 AA.
AC P98200;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE POTENTIAL PHOSPHOLIPID-TRANSPORTING ATPASE IB (EC 3.6.1.-).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ICR;
RX MEDLINE-20473714; PubMed-11015572;
RA Halleck M.S., Lawler J.F. Jr., Blackshaw S., Gao L., Nagarajan P.,
RA Hacker C., Pyle S., Newman J.T., Nakanishi Y., Ando H., Weinstein D.,
RA Williamson P., Schlegel R.A.;
RT "Differential expression of putative transbilayer amphipath
RT transporters";
RL Physiol. Genomics 1:139-150(1999).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: FOUND IN TESTIS, HEART AND BRAIN. MOST
CC ABUNDANT IN TESTIS. ALSO DETECTED IN FETAL TISSUES.

FT DOMAIN 131 136 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 137 156 POTENTIAL.
 FT DOMAIN 157 340 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 341 362 POTENTIAL.
 FT DOMAIN 363 389 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 390 411 POTENTIAL.
 FT DOMAIN 412 949 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 950 970 POTENTIAL.
 FT DOMAIN 971 982 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 983 1002 POTENTIAL.
 FT DOMAIN 1003 1032 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1033 1054 POTENTIAL.
 FT DOMAIN 1055 1068 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1069 1091 POTENTIAL.
 FT DOMAIN 1092 1097 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1098 1118 POTENTIAL.
 FT DOMAIN 1119 1138 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1139 1163 POTENTIAL.
 FT DOMAIN 1164 1251 CYTOPLASMIC (POTENTIAL).
 FT MOD.RES 454 454 PHOSPHORYLATION (BY SIMILARITY).
 FT METAL 893 893 MAGNESIUM (BY SIMILARITY).
 FT METAL 897 897 MAGNESIUM (BY SIMILARITY).
 FT VARIANT 288 288 L -> S (IN PFIC1).
 FT VARIANT 308 308 /FTIG-VAR_008809.
 FT VARIANT 645 699 /FTIG-VAR_008810.
 FT VARIANT 661 661 MISSING (IN PFIC1).
 FT VARIANT 892 892 /FTIG-VAR_008811.
 FT VARIANT 795 797 I -> T (IN BRIC).
 FT VARIANT 1251 143725 MW; 271EPE24EDA6E144 CRC64; G -> V (IN PFIC1).
 FT SEQUENCE 1251 AA; 143725 MW; 271EPE24EDA6E144 CRC64; /FTIG-VAR_008813.
 FT SEQUENCE 1251 AA; 143725 MW; 271EPE24EDA6E144 CRC64; MISSING (IN BRIC).
 FT SEQUENCE 1251 AA; 143725 MW; 271EPE24EDA6E144 CRC64; /FTIG-VAR_008814.

Query Match 44.8%; Score 43; DB 1; Length 1251;
 Best Local Similarity 63.6%; Pred. No. 1.5e+02;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 9 AVALAVGDSN 19
 I: |||:|:|:
 DB 886 AYLAIGDGN 896

RESULT 13
 YJBE_ECOLI STANDARD; PRT; 80 AA.
 AC P32686; Q923E1;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DE 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HYPOHETICAL 7.4 KDA PROTEIN IN PGI-XYLE INTERGENIC REGION PRECURSOR.
 GN YJBE OR B4026.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MGL1655;
 RX MEDLINE=94089392; PubMed=8265357;
 RA Blattner F.R., Burland V.D., Plunkett G. III; Sofia H.J.,
 RA Daniels D.L.;
 RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
 RT region from 89.2 to 92.8 minutes."
 RL Nucleic Acids Res. 21:5408-5417(1993).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U00006; AAC43120.1; -;
 DR EMBL; AE000476; AAC76996.1; -;
 DR EcoGene; EGI1923; YJBE.
 KW Hypothetical protein; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 80 HYPOHETICAL PROTEIN YJBE.
 SQ SEQUENCE 80 AA; 7386 MW; A34F87939E466272 CRC64;
 Query Match 44.3%; Score 42.5; DB 1; Length 80;
 Best Local Similarity 61.1%; Pred. No. 19;
 Matches 11; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

OY 7 ALAVALAV---GDGSNIG 21
 I: ||||| ||||| I
 DB 53 AVGVAAATGGGDSNTG 70

RESULT 14
 PBP_MANSE STANDARD; PRT; 168 AA.
 ID PBP_MANSE
 AC P18959;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PHEROMONE-BINDING PROTEIN PRECURSOR (PBP).
 OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
 OC Spingilodea; Spingilidae; Spinginae; Manduca.
 OX NCBI_TaxID=7130;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-56.
 RX MEDLINE=89071794; PubMed=3200861;
 RX Gyorgyi T.K., Roby-Shemkovitz A.J., Lerner M.R.;
 RT "Characterization and cDNA cloning of the pheromone-binding protein
 RT from the tobacco hornworm, Manduca sexta: a tissue-specific
 RT developmentally regulated protein."
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9851-9855(1988).
 RN [2]
 RP SEQUENCE OF 27-61.
 RX MEDLINE=91186129; PubMed=2010751;
 RX Vogt R.G., Prestwich G.D., Lerner M.R.;
 RT "Odorant-binding-protein subfamilies associate with distinct classes
 RT of olfactory receptor neurons in insects."
 RL J. Neurobiol. 22:74-84(1991).
 CC -!- FUNCTION: THIS MAJOR SOLUBLE PROTEIN IN OLFACTORY SENSILLA OF MALE
 CC MOTHS MIGHT SERVE TO SOLUBILIZE THE EXTREMELY HYDROPHOBIC
 CC PHEROMONE MOLECULES AND TO TRANSPORT PHEROMONE THROUGH THE AQUEOUS
 CC LYMPH TO RECEPTORS LOCATED ON OLFACTORY CILIA. PBP IS ALSO FOUND
 CC IN SENSILLA FROM FEMALE M. SEXTA ANTENNAE.
 CC -!- SUBUNIT: HOMODIMER (PROBABLE).
 CC -!- TISSUE SPECIFICITY: ANTENNA.
 CC -!- DEVELOPMENTAL STAGE: ITS SYNTHESIS OCCURS AROUND THE TIME OF
 CC ECLOSION.
 CC -!- SIMILARITY: BELONGS TO THE PBP/GOBP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M21797; AAA29325.1; -;
 DR EMBL; M21798; AAA29326.1; -;
 DR PIR; A31770; A31770.
 DR InterPro; IPR000746; -;

DR Pfam: PF01395; PBP_GOBP; 1.
KW PRINTS; PR00484; PBPGOBP.
DR Phormone response; Transport; Signal.
FT SIGNAL 1 26
FT CHAIN 27 168 PHORMONE-BINDING PROTEIN.
FT VARIANT 29 29 D -> E.
FT VARIANT 120 120 I -> V.
SQ SEQUENCE 168 AA; 18516 MW; 0FC1F18D1908ADF1 CRC64;

Query Match 43.8%; Score 42; DB 1; Length 168;
Best Local Similarity 56.2%; Pred. No. 40;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 MRVLVLALVALAVG 16
I: I: I: I: I: I:
Db 8 MKVAVVAIVVYLAAGN 23

RESULT 15
APAL_SALSA
ID APAL_SALSA STANDARD; PRT; 258 AA.
AC P27007:
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE APOLIPOPROTEIN A-I PRECURSOR (APO-AI).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92009208; PubMed=1916288;
RA Powell R., Higgins D.G., Wolff J., Byrnes L., Stack M., Sharp P.M.,
RA Gannon F.:
RT "The salmon gene encoding apolipoprotein A-I: cDNA sequence, tissue
RT expression and evolution.";
RL Gene 104:155-161(1991).
CC -!- FUNCTION: APOA-1 PARTICIPATES IN THE REVERSE TRANSPORT OF
CC CHOLESTEROL FROM TISSUES TO THE LIVER FOR EXCRETION BY PROMOTING
CC CHOLESTEROL EFFLUX FROM TISSUES AND BY ACTING AS A COFACTOR FOR
CC THE LECITHIN CHOLESTEROL ACYLTRANSFERASE (LCAT).
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- TISSUE SPECIFICITY: MAJOR PROTEIN OF PLASMA HDL, ALSO FOUND IN
CC CHYLOMICRONS. EXPRESSED IN LIVER, INTESTINE, AND MUSCLE.
CC -!- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52237; CAA36482.1; -
DR PIR: JH0472; JH0472.
DR PIR: S26810; S26810.
DR InterPro: IPR000074; -
DR Pfam: PF01442; Apolipoprotein; 1.
KW Plasma; Lipid transport; HDL; Cholesterol metabolism; Repeat; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 31 POTENTIAL.
FT CHAIN 32 258 APOLIPOPROTEIN A-I.
FT DOMAIN 32 63 3 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 64 258 10 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 64 85 1.
FT REPEAT 86 106 2.
FT REPEAT 107 117 3 (HALF-LENGTH).
FT REPEAT 118 139 4.

FT REPEAT 140 161 5.
FT REPEAT 162 183 6.
FT REPEAT 184 205 7.
FT REPEAT 206 227 8.
FT REPEAT 228 238 9.
FT REPEAT 239 258 10.
SQ SEQUENCE 258 AA; 29448 MW; 4AB6C7E82D8179F3 CRC64;

Query Match 43.8%; Score 42; DB 1; Length 258;
Best Local Similarity 60.0%; Pred. No. 57;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 MRVLVLALVALAVG 15
I: I: I: I: I: I:
Db 1 MKFLVLALTILLAAG 15

Search completed: May 3, 2001, 15:52:54
Job time: 73 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2001, 15:51:41 ; Search time 18.59 Seconds
(without alignments)
64.574 Million cell updates/sec

Title: US-09-426-776-10
Perfect score: 96
Sequence: 1 MRVLVLAVALVAGDSNLG 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0401.*
1: /SID6/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID6/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID6/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID6/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID6/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID6/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID6/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID6/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID6/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID6/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID6/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID6/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID6/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID6/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID6/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID6/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID6/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID6/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID6/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID6/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID6/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID6/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88	91.7	21	Y92778	O. aureus vitellog
2	88	91.7	34	Y92780	Vtgs-CAT fusion p
3	88	91.7	38	Y92782	Vtgs-EGFP fusion
4	88	91.7	51	Y92779	Vtgs-CrFCS fusio
5	88	91.7	51	Y92783	Vtgs-beta-lactama
6	45	46.9	313	G08134	Arabidopsis thalia
7	45	46.9	313	G14879	Arabidopsis thalia
8	45	46.9	313	G14879	Arabidopsis thalia
9	45	46.9	313	G14879	Arabidopsis thalia
10	45	46.9	325	G08133	Arabidopsis thalia
11	45	46.9	325	G14878	Arabidopsis thalia

12	45	46.9	325	21	G49240	Arabidopsis thalia
13	45	46.9	325	21	G49245	Arabidopsis thalia
14	45	46.9	345	21	G49244	Arabidopsis thalia
15	45	46.9	349	21	G08132	Arabidopsis thalia
16	45	46.9	353	21	G14877	Arabidopsis thalia
17	45	46.9	353	21	G49239	Arabidopsis thalia
18	45	46.9	908	21	B42511	Human ORFX ORF2275
19	44	45.8	381	18	W35283	Human acid sphingo
20	44	45.8	628	14	R30644	DeltaR608 ASM. Ho
21	44	45.8	628	18	W35262	ASM protein. Homo
22	44	45.8	629	14	R30642	R496L ASM. Homo s
23	44	45.8	629	14	R30643	L302P ASM. Homo s
24	44	45.8	629	14	R30645	Human acid sphingo
25	44	45.8	629	18	W35260	Human acid sphingo
26	44	45.8	629	18	W35261	Human acid sphingo
27	44	45.8	629	18	W35263	Human acid sphingo
28	44	45.8	1000	21	B41414	Human ORFX ORF1178
29	44	45.8	1027	14	R42203	Protein L. Peptoc
30	44	45.8	1027	14	R43699	Protein L. Peptoc
31	43	44.8	222	17	W00679	Beta-casein. Capr
32	43	44.8	423	21	Y45093	Mouse lymphoid der
33	43	44.8	466	21	Y95728	Cosmid CHRIM5 enco
34	43	44.8	683	14	R40386	betaIG-H3 protein.
35	43	44.8	683	16	R80573	Human beta-IG-H3 (
36	43	44.8	683	21	B11897	Human colon tumour
37	43	44.8	731	21	B41613	Human ORFX ORF1377
38	43	44.8	887	19	W77409	Mycobacterium Kans
39	43	44.8	1251	20	Y06516	Human cholestatin.
40	42.5	44.3	186	14	R32438	Wheat germ aggluti
41	41	42.7	157	21	G54665	Arabidopsis thalia
42	41	42.7	188	21	G18321	Arabidopsis thalia
43	41	42.7	225	14	R32019	Beta-casein. Homo
44	41	42.7	225	14	R33451	Sequence encoded b
45	41	42.7	225	19	W53664	Human beta-casein.

ALIGNMENTS

RESULT 1
Y92778 Y92778 standard; Peptide; 21 AA.
XX
AC Y92778;
XX
DT 29-ARG-2000 (first entry)
XX
DE O. aureus vitellogenin secretory signal sequence.
XX
XX Vtgs; vitellogenin; secretory signal sequence; gene expression;
KW oestrogen receptor binding protein; systemic circulation.
XX
XX Oreochromis aureus.
XX
XX Key Location/Qualifiers
FH Cleavage-site 15..16.
XX
XX
PN W0200026366-A1.
XX
PD 11-MAY-2000.
XX
PF 29-OCT-1999; 99WO-SG00108.
XX
PR 30-OCT-1998; 98US-0106426.
PR 26-OCT-1999; 99US-0426776.
XX
XX (UYSI-) UNIV SINGAPORE NAT.
PA (LAMT/) LAM T J.
XX
PI Ding JL, Tan NS, Ho B;
XX WPI; 2000-365615/31.
DR N-PSDB; A28495.

XX Isolated nucleic acid for assaying for heterologous gene expression,
 PT detecting presence of compound that binds to estrogen receptor or
 PT producing desired protein from host cell comprises nucleotide sequence
 PT encoding secretory signal sequence
 XX
 PS Claim 1; Page 36; 73pp; English.
 XX
 CC This sequence is that of a piscine, Oreochromis aureus, vitellogenin
 CC secretory sequence (Vtgss). This and variants that comprise conservative
 CC replacements that retain the biological activities of directing secretion
 CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Vtgss can be
 CC fused to either a reporter protein or a lipopolysaccharide-binding
 CC protein coding sequence. The isolated nucleic acid is useful for assaying
 CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an oestrogen receptor in a sample or producing a desired
 CC protein from a host cell. It can also be used in a method for obtaining
 CC systemic circulation of a desired protein in a transgenic or chimeric
 CC host organism.
 XX
 SQ Sequence 21 AA;

Query Match 91.7%; Score 88; DB 21; Length 21;
 Best Local Similarity 95.2%; Pred. NO. 3.4e-07;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MRVLVALAVAGDGSNLG 21
 |||||
 Db 1 mrvlvalavagdqsnlg 21

RESULT 2

Y92780
 ID Y92780 standard; Protein; 34 AA.

XX Y92780;

XX 29-AUG-2000 (first entry)

XX Vtgss-CAT fusion protein (partial).

XX Vtgss: vitellogenin; secretory signal sequence; gene expression;
 KW oestrogen receptor binding protein; systemic circulation; CAT.

XX Chimeric - Oreochromis aureus.

OS Chimeric - Synthetic.

XX Key Location/Qualifiers
 FT Peptide 1..21

FT /label= secretory_signal_sequence
 FT /note= "O. aureus"

FT Cleavage-site 15..16

FT Protein 22..34

FT /note= "CAT N-terminal"

XX WO200026366-A1.

XX 11-MAY-2000.

XX 29-OCT-1999; 99WO-SG00108.

XX 30-OCT-1998; 98US-0106426.

XX 26-OCT-1999; 99US-0426776.

XX (UYSI-) UNIV SINGAPORE NAT.

XX (LAMT/) LAM T J.

XX Ding JL, Tan NS, Ho B;

XX WPI; 2000-365615/31.

DR N-PSDB; A28499.

XX Isolated nucleic acid for assaying for heterologous gene expression,
 PT detecting presence of compound that binds to estrogen receptor or
 PT producing desired protein from host cell comprises nucleotide sequence
 PT encoding secretory signal sequence
 XX
 PS Example 2; Fig 5B; 73pp; English.
 XX
 CC A reporter CAT system that uses the piscine, Oreochromis aureus,
 CC vitellogenin secretory sequence (Vtgss), designated psp-VtgCAT was
 CC constructed. A28499 and A28722 comprise the 5'- and 3'-ends of the
 CC construct insert. Vtgss and variants that comprise conservative
 CC replacements that retain the biological activities of directing secretion
 CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Vtgss can be
 CC fused to either a reporter protein or a lipopolysaccharide-binding
 CC protein coding sequence. The isolated nucleic acid is useful for assaying
 CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an oestrogen receptor in a sample or producing a desired
 CC protein from a host cell. It can also be used in a method for obtaining
 CC systemic circulation of a desired protein in a transgenic or chimeric
 CC host organism.
 XX
 SQ Sequence 34 AA;

Query Match 91.7%; Score 88; DB 21; Length 34;
 Best Local Similarity 95.2%; Pred. NO. 5.8e-07;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MRVLVALAVAGDGSNLG 21
 |||||
 Db 1 mrvlvalavagdqsnlg 21

RESULT 3

Y92782

ID Y92782 standard; Protein; 38 AA.

XX Y92782;

XX 29-AUG-2000 (first entry)

XX Vtgss-EGFP fusion protein (partial).

XX Vtgss: vitellogenin; secretory signal sequence; gene expression;
 KW oestrogen receptor binding protein; systemic circulation; EGFP.

XX Chimeric - Oreochromis aureus.

OS Chimeric - Synthetic.

XX Key Location/Qualifiers
 FT Peptide 1..21

FT /label= Vtgss

FT Cleavage-site 15..16

FT Protein 28..38

FT /label= EGFP

XX WO200026366-A1.

XX 11-MAY-2000.

XX 29-OCT-1999; 99WO-SG00108.

XX 30-OCT-1998; 98US-0106426.

XX 26-OCT-1999; 99US-0426776.

XX (UYSI-) UNIV SINGAPORE NAT.

XX (LAMT/) LAM T J.

XX Ding JL, Tan NS, Ho B;

XX WPI; 2000-365615/31.

DR

DR N-PSDB; A28502.

XX Isolated nucleic acid for assaying for heterologous gene expression,
 PT detecting presence of compound that binds to estrogen receptor or
 PT producing desired protein from host cell comprises nucleotide sequence
 PT encoding secretory signal sequence

PS Example 3; Fig 8B; 73pp; English.

XX

CC A reporter GFP system that uses the piscine, Oreochromis aureus,
 CC vitellogenin secretory sequence (Vtgss), designated pVtEGFP was
 CC constructed. Vtgss and variants that comprise conservative
 CC replacements that retain the biological activities of directing secretion
 CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Vtgss can be
 CC fused to either a reporter protein or a lipopolysaccharide-binding
 CC protein coding sequence. The isolated nucleic acid is useful for assaying
 CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an estrogen receptor in a sample or producing a desired
 CC protein from a host cell. It can also be used in a method for obtaining
 CC systemic circulation of a desired protein in a transgenic or chimeric
 CC host organism.

XX

SQ Sequence 38 AA;

Query Match 91.7%; Score 88; DB 21; Length 38;
 Best Local Similarity 95.2%; Pred. No. 6.5e-07;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRVLVLALVALAVGDGNSLNG 21
 |||||
 Db 1 mrvlvalavalavgdgnsng 21

RESULT 4

Y92779

ID Y92779 standard; Protein; 51 AA.

XX

AC Y92779;

XX

XX 29-AUG-2000 (first entry)

DT

DE Vtgss-CrFCES fusion protein.

XX

XX Vtgss; vitellogenin; secretory signal sequence; gene expression;
 KW oestrogen receptor binding protein; systemic circulation; factor C.

KW

OS Chimeric - Oreochromis aureus.

OS Chimeric - Carcinoscopus rotundicauda.

XX

FH Key Location/Qualifiers

FT Peptide 1..21

FT /label= secretory_signal_sequence

FT /note= "from O. aureus"

FT Cleavage-site 15..16

FT Domain 22..51

FT /label= LPS-binding_domain

FT /note= "from C. rotundicauda Factor C"

XX

XX WO200026366-A1.

XX

XX 11-MAY-2000.

XX

XX 29-OCT-1999; 99WO-SG00108.

XX

XX 30-OCT-1998; 98US-0106426.

PR 26-OCT-1999; 99US-0426776.

XX

XX (UYSI-) UNIV SINGAPORE NAT.

PA (LAM T J.

XX

PI Ding JL, Tan NS, Ho B;

XX WPI; 2000-365615/31.

DR N-PSDB; A28496.

XX

XX Isolated nucleic acid for assaying for heterologous gene expression,
 PT detecting presence of compound that binds to estrogen receptor or
 PT producing desired protein from host cell comprises nucleotide sequence
 PT encoding secretory signal sequence

XX

XX Example 1; Fig 2A; 73pp; English.

PS

XX The piscine, Oreochromis aureus, vitellogenin secretory sequence (Vtgss)
 CC was fused upstream to the EcoRI-SalI cDNA fragment encoding the
 CC lipopolysaccharide-binding domain of Carcinoscopus rotundicauda Factor
 CC C CrFCES for expression and secretion of recombinant ES protein from
 CC Drosophila cells. Vtgss and variants that comprise conservative
 CC replacements that retain the biological activities of directing secretion
 CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Vtgss can be
 CC fused to either a reporter protein or a lipopolysaccharide-binding
 CC protein coding sequence. The isolated nucleic acid is useful for assaying
 CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an estrogen receptor in a sample or producing a desired
 CC protein from a host cell. It can also be used in a method for obtaining
 CC systemic circulation of a desired protein in a transgenic or chimeric
 CC host organism.

XX

SQ Sequence 51 AA;

Query Match 91.7%; Score 88; DB 21; Length 51;
 Best Local Similarity 95.2%; Pred. No. 9e-07;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRVLVLALVALAVGDGNSLNG 21
 |||||
 Db 1 mrvlvalavalavgdgnsng 21

RESULT 5

Y92783

ID Y92783 standard; Protein; 51 AA.

XX

AC Y92783;

XX

XX 29-AUG-2000 (first entry)

DT

DE Vtgss-beta-lactamase fusion protein (partial).

XX

XX Vtgss; vitellogenin; secretory signal sequence; gene expression;
 KW oestrogen receptor binding protein; systemic circulation;
 KW beta-lactamase.

KW

OS Chimeric - Oreochromis aureus.

OS Chimeric - Synthetic.

XX

FH Key Location/Qualifiers

FT Peptide 1..21

FT /label= Vtgss

FT Cleavage-site 15..16

FT Protein 22..51

FT /note= "beta-lactamase mature protein"

XX

XX WO200026366-A1.

XX

XX 11-MAY-2000.

XX

XX 29-OCT-1999; 99WO-SG00108.

XX

XX 30-OCT-1998; 98US-0106426.

PR 26-OCT-1999; 99US-0426776.

XX

XX (UYSI-) UNIV SINGAPORE NAT.

PA

PA (LAWT/) LAM T J.
XX Ding JL, Tan NS, Ho B;
PI WPI: 2000-365615/31.
XX N-PSDB; A28507.
DR
XX Isolated nucleic acid for assaying for heterologous gene expression,
PT detecting presence of compound that binds to estrogen receptor or
PT producing desired protein from host cell comprises nucleotide sequence
PT encoding secretory signal sequence
XX
XX Example 6; Fig 14A; 73pp; English.
PS
XX A reporter beta-lactamase system that uses the *Oreochromis aureus*,
CC vitellogenin secretory sequence (Vtgss), designated pBADVtgblactKana was
CC constructed. Vtgss and variants that comprise conservative
CC replacements that retain the biological activities of directing secretion
CC of a fusion protein from a cell and cleavage of the secretory signal
CC sequence from the fusion protein, are new. DNA encoding the Vtgss can be
CC fused to either a reporter protein or a lipopolysaccharide-binding
CC protein coding sequence. The isolated nucleic acid is useful for assaying
CC for heterologous gene expression, detecting the presence of a compound
CC that binds to an estrogen receptor in a sample or producing a desired
CC protein from a host cell. It can also be used in a method for obtaining
CC systemic circulation of a desired protein in a transgenic or chimeric
CC host organism.
XX
XX Sequence 51 AA;
SQ

Query Match 91.7%; Score 88; DB 21; Length 51;
Best Local Similarity 95.2%; Pred. No. 9e-07;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MRVLVLAVALAVGDCSNL 21
| | | | | | | | | | | | | | | | | | | | | |
Db 1 mrvlvialavalavgdqsnlg 21

RESULT 6
G08134
ID G08134 standard; Protein; 313 AA.
XX
AC G08134;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 5552.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR

21-APR-1999; 99US-0130449.
23-APR-1999; 99US-0130510.
28-APR-1999; 99US-0130891.
30-APR-1999; 99US-0131449.
30-APR-1999; 99US-0132048.
30-APR-1999; 99US-0132407.
04-MAY-1999; 99US-0132484.
05-MAY-1999; 99US-0132485.
06-MAY-1999; 99US-0132486.
07-MAY-1999; 99US-0132487.
11-MAY-1999; 99US-0132863.
14-MAY-1999; 99US-0134256.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
18-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
02-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.

PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 46.9%; Score 45; DB 21; Length 313;
 Best Local Similarity 56.2%; Pred. No. 25;
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 LVLALAVAVGDSN 19
 I:: |||::|||
 Db 133 llimaavillsvegsn 148

RESULT 7
 GI4879
 ID GI4879 standard; Protein; 313 AA.
 XX GI4879;
 AC GI4879;
 DT 17-OCT-2000 (first entry)
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 14907.
 DE Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 XX EP1033405-A2.
 PD 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-0301439.
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.

Protein identification; signal transduction pathway; metabolic pathway;
 hybridisation assay; genetic mapping; gene expression control; promoter;
 termination sequence.

PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-01334256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 19-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 24-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136392.
PR 27-MAY-1999; 99US-0136782.
PR 28-MAY-1999; 99US-0137222.
PR 01-JUN-1999; 99US-0137528.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144003.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154773.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.

```
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 23-OCT-1999; 99US-0161405.
PR 23-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 46.9%; Score 45; DB 21; Length 313;
Best Local Similarity 56.2%; Pred. No. 25;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 LVLALAVAVGDSN 19
   I: I I I I I I I I
Db 133 lllmaavllsvgegsn 148

RESULT 8
G49241
ID G49241 standard; Protein; 313 AA.
XX
AC G49241;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 62272.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX.
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
```

```
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.
PR 29-OCT-1999; 99US-0162144.

Query Match 46.9%; Score 45; DB 21; Length 313;
Best Local Similarity 56.2%; Pred. No. 25;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 LVLALAVAVGDSN 19
Db 133 lllmaavllsvgegsn 148
:: |||::|||

RESULT 9
G49246
ID G49246 standard; Protein; 313 AA.
XX AC G49246;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 62279.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EPI033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
```

PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 24-JUN-1999; 99US-0140823.
 PR 28-JUN-1999; 99US-0140931.
 PR 29-JUN-1999; 99US-0141287.
 PR 30-JUN-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 01-JUL-1999; 99US-0142055.
 PR 02-JUL-1999; 99US-0142390.
 PR 06-JUL-1999; 99US-0142803.
 PR 08-JUL-1999; 99US-0142920.
 PR 09-JUL-1999; 99US-0142977.
 PR 12-JUL-1999; 99US-0143542.
 PR 13-JUL-1999; 99US-0143624.
 PR 14-JUL-1999; 99US-0144005.
 PR 15-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 16-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 06-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 30-AUG-1999; 99US-0151080.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0155139.
 PR 22-SEP-1999; 99US-0155179.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.

```
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match          .46.9%; Score 45; DB 21; Length 313;
Best Local Similarity 56.2%; Pred. No. 25;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 LVLAVALAVAGDGSN 19
Db 133 lliimaavllsvgegn 148
      |:: || |::|||
RESULT 10
G08133
ID G08133 standard; Protein; 325 AA.
XX
AC G08133;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 5551.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0123788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
```

```
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156438.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match          46.9%; Score 45; DB 21; Length 325;
Best Local Similarity 56.2%; Pred. No. 26;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 LVLALAVAVGDSN 19
Db 145 llimaavllsvgegn 160
l:: |||::|||
145 llimaavllsvgegn 160

RESULT 11
GI4878
ID GI4878 standard; Protein; 325 AA.
XX
AC GI4878;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 14906.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
```

us-09-426-776-10.rag

PR	16-AUG-1999;	99US-01493168;
PR	17-AUG-1999;	99US-01493175;
PR	18-AUG-1999;	99US-01493261;
PR	20-AUG-1999;	99US-01497222;
PR	20-AUG-1999;	99US-01497233;
PR	20-AUG-1999;	99US-01497233;
PR	20-AUG-1999;	99US-01499229;
PR	23-AUG-1999;	99US-01499202;
PR	23-AUG-1999;	99US-01499300;
PR	25-AUG-1999;	99US-01505666;
PR	26-AUG-1999;	99US-01508884;
PR	27-AUG-1999;	99US-01510665;
PR	27-AUG-1999;	99US-01510666;
PR	27-AUG-1999;	99US-01510803;
PR	30-AUG-1999;	99US-01513033;
PR	30-AUG-1999;	99US-01513033;
PR	31-AUG-1999;	99US-01514338;
PR	01-SEP-1999;	99US-01519330;
PR	07-SEP-1999;	99US-01523663;
PR	10-SEP-1999;	99US-01530770;
PR	13-SEP-1999;	99US-01530758;
PR	15-SEP-1999;	99US-01540181;
PR	16-SEP-1999;	99US-01540339;
PR	20-SEP-1999;	99US-01547779;
PR	22-SEP-1999;	99US-01551339;
PR	23-SEP-1999;	99US-01554886;
PR	24-SEP-1999;	99US-01556559;
PR	28-SEP-1999;	99US-01564558;
PR	29-SEP-1999;	99US-01565966;
PR	04-OCT-1999;	99US-01571177;
PR	05-OCT-1999;	99US-01577553;
PR	06-OCT-1999;	99US-01578665;
PR	07-OCT-1999;	99US-01580229;
PR	08-OCT-1999;	99US-01582322;
PR	12-OCT-1999;	99US-01583669;
PR	13-OCT-1999;	99US-01592933;
PR	13-OCT-1999;	99US-01592934;
PR	13-OCT-1999;	99US-01592935;
PR	14-OCT-1999;	99US-01593329;
PR	14-OCT-1999;	99US-01593329;
PR	14-OCT-1999;	99US-01593331;
PR	14-OCT-1999;	99US-01596337;
PR	14-OCT-1999;	99US-01596338;
PR	18-OCT-1999;	99US-01595894;
PR	21-OCT-1999;	99US-01607411;
PR	21-OCT-1999;	99US-01607677;
PR	21-OCT-1999;	99US-01607688;
PR	21-OCT-1999;	99US-01607770;
PR	21-OCT-1999;	99US-01608114;
PR	21-OCT-1999;	99US-01608115;
PR	22-OCT-1999;	99US-01608980;
PR	22-OCT-1999;	99US-01609081;
PR	22-OCT-1999;	99US-01609089;
PR	25-OCT-1999;	99US-01614004;
PR	25-OCT-1999;	99US-01614005;
PR	25-OCT-1999;	99US-01614066;
PR	26-OCT-1999;	99US-01613559;
PR	26-OCT-1999;	99US-01613600;
PR	26-OCT-1999;	99US-01613611;
PR	28-OCT-1999;	99US-01619200;
PR	28-OCT-1999;	99US-01619922;
PR	28-OCT-1999;	99US-01619933;
PR	29-OCT-1999;	99US-01621242;

```
QY 4 LVLALAVALAVGDGSN 19
    I:: I I I:: I I I I
Db 145 lliinaavllsvgegsn 160

RESULT 12
G49240
```


ID G49240 standard; Protein; 325 AA.
 AC G49240;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 62271.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 07-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 19-MAY-1999; 99US-0134768.
 PR 20-MAY-1999; 99US-0134941.
 PR 21-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 24-JUN-1999; 99US-0140823.
 PR 28-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147316.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.

```
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 46.9%; Score 45; DB 21; Length 325;
Best Local Similarity 56.2%; Pred. No. 26;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 LVLALAVAVGDSN 19
Db 145 llimaavllsvgegn 160
::: |||:|||||

RESULT 13
G49245
ID G49245 standard; Protein; 325 AA.
XX
AC G49245;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 62278.

XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
```

PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144009.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 46.9%; Score 45; DB 21; Length 325;

Best Local Similarity 56.2%; Pred. No. 26;

Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 LVLAVALAVGDSN 19

Db 145 llimaavllsvgegsn 160

RESULT 14

G49244

ID G49244 standard; Protein: 345 AA.

XX AC G49244;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 62277.

XX KW Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX OS Arabidopsis thaliana.

PN	EP1033405-A2.	PR	01-JUL-1999;	99US-0142154.
XX		PR	02-JUL-1999;	99US-0142055.
PD		PR	06-JUL-1999;	99US-0142390.
XX	06-SEP-2000.	PR	08-JUL-1999;	99US-0142803.
PF		PR	09-JUL-1999;	99US-0142920.
XX	25-FEB-2000; 2000EP-0301439.	PR	12-JUL-1999;	99US-0142977.
XX		PR	13-JUL-1999;	99US-0143542.
PR	25-FEB-1999;	PR	14-JUL-1999;	99US-0143624.
PR	05-MAR-1999;	PR	15-JUL-1999;	99US-0144005.
PR	09-MAR-1999;	PR	16-JUL-1999;	99US-0144085.
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144086.
PR	25-MAR-1999;	PR	19-JUL-1999;	99US-0144325.
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144331.
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144332.
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144333.
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144334.
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144335.
PR	19-APR-1999;	PR	20-JUL-1999;	99US-0144352.
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145088.
PR	30-APR-1999;	PR	22-JUL-1999;	99US-0145085.
PR	04-MAY-1999;	PR	22-JUL-1999;	99US-0145087.
PR	03-MAY-1999;	PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	PR	22-JUL-1999;	99US-0145145.
PR	07-MAY-1999;	PR	23-JUL-1999;	99US-0145218.
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145918.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145919.
PR	18-MAY-1999;	PR	28-JUL-1999;	99US-0145951.
PR	19-MAY-1999;	PR	02-AUG-1999;	99US-0146386.
PR	20-MAY-1999;	PR	02-AUG-1999;	99US-0146388.
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146389.
PR	24-MAY-1999;	PR	03-AUG-1999;	99US-0147038.
PR	25-MAY-1999;	PR	03-AUG-1999;	99US-0147204.
PR	27-MAY-1999;	PR	04-AUG-1999;	99US-0147302.
PR	28-MAY-1999;	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	PR	09-AUG-1999;	99US-0147493.
PR	08-JUN-1999;	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	PR	09-AUG-1999;	99US-0148171.
PR	10-JUN-1999;	PR	10-AUG-1999;	99US-0148171.
PR	10-JUN-1999;	PR	11-AUG-1999;	99US-0148319.
PR	14-JUN-1999;	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	PR	13-AUG-1999;	99US-0148565.
PR	16-JUN-1999;	PR	13-AUG-1999;	99US-0148684.
PR	17-JUN-1999;	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	PR	25-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	PR	27-AUG-1999;	99US-0151080.
PR	21-JUN-1999;	PR	30-AUG-1999;	99US-0151303.
PR	22-JUN-1999;	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	PR	31-AUG-1999;	99US-0151930.
PR	23-JUN-1999;	PR	07-SEP-1999;	99US-0152363.
PR	24-JUN-1999;	PR	10-SEP-1999;	99US-0153070.
PR	24-JUN-1999;	PR	10-SEP-1999;	99US-0153758.
PR	29-JUN-1999;	PR	13-SEP-1999;	99US-0154018.
PR	30-JUN-1999;	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	PR	16-SEP-1999;	99US-0154039.

```
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 46.9%; Score 45; DB 21; Length 345;
Best Local Similarity 56.2%; Pred. No. 27;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 LVLALAVLAVGDSN 19
   |::|||::|||
Db 165 lllmaavllsvgegn 180

RESULT 15
G08132
ID G08132 standard; Protein; 349 AA.
XX
AC G08132;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 5550.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135253.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
```

PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 46.9%; Score 45; DB 21; Length 349;
Best Local Similarity 56.2%; Pred. No. 28;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Oy 4 LVLALAVLAYGDGNS 19
|:: || |::|||
Db 169 llimaavllsvgegn 184

Search completed: May 3, 2001, 15:52:21
Job time: 40 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: May 3, 2001, 15:51:41 ; Search time 11.61 Seconds
(without alignments)
34.748 Million cell updates/sec

Title: US-09-426-776-10
Perfect score: 96
Sequence: 1 MRVLALAVALLAVGDSNLG 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues
Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/6C_COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	45.8	629	1	US-08-250-740-33
2	44	45.8	629	1	US-07-695-472B-2
3	44	45.8	1027	4	US-08-446-137B-2
4	43	44.8	222	2	US-08-391-743A-2
5	43	44.8	683	1	US-07-878-960-2
6	43	44.8	683	2	US-08-477-396A-17
7	40	41.7	361	3	US-09-120-365-75
8	40	41.7	516	2	US-08-676-166A-2
9	40	41.7	777	2	US-08-477-396A-4
10	40	41.7	1784	4	US-09-040-738-2
11	39	40.6	455	2	US-08-738-172-4
12	39	40.6	507	1	US-08-484-493-12
13	39	40.6	507	1	US-08-484-494-12
14	39	40.6	507	2	US-08-345-212-12
15	39	40.6	507	4	US-09-249-003-12
16	39	40.6	509	1	US-08-445-586-8
17	38	39.6	90	4	US-09-124-671-25
18	38	39.6	100	1	US-08-384-367-2
19	38	39.6	109	4	US-09-124-671-23
20	38	39.6	109	4	US-09-124-671-27
21	38	39.6	109	4	US-09-124-671-29
22	38	39.6	374	1	US-08-095-726-14
23	38	39.6	374	1	US-08-096-623A-14
24	38	39.6	1241	4	US-09-040-774-2
25	37	38.5	214	1	US-07-953-230A-11
26	37	38.5	402	3	US-08-948-997-4
27	37	38.5	402	4	US-09-348-817A-4

28	37	38.5	521	2	US-08-406-855A-19	Sequence 19, Appl
29	37	38.5	521	3	US-09-206-899-19	Sequence 20, Appl
30	37	38.5	559	2	US-08-406-855A-20	Sequence 20, Appl
31	37	38.5	559	3	US-09-206-899-20	Sequence 18, Appl
32	36.5	38.0	63	3	US-09-061-026-18	Sequence 18, Appl
33	36.5	38.0	63	4	US-09-466-138-18	Sequence 36, Appl
34	36	37.5	148	2	US-08-888-497-36	Sequence 36, Appl
35	36	37.5	148	5	PCT-US94-07926-36	Sequence 36, Appl
36	36	37.5	423	3	US-08-494-907-14	Sequence 14, Appl
37	36	37.5	423	5	PCT-US96-10986-14	Sequence 14, Appl
38	36	37.5	474	1	US-08-222-619-5	Sequence 5, Appl
39	36	37.5	474	5	PCT-US95-04075-5	Sequence 5, Appl
40	36	37.5	488	1	US-08-243-542-1	Sequence 1, Appl
41	36	37.5	488	1	US-08-477-407-1	Sequence 1, Appl
42	36	37.5	488	1	US-08-484-355-1	Sequence 1, Appl
43	36	37.5	524	1	US-08-243-542-2	Sequence 2, Appl
44	36	37.5	524	1	US-08-477-407-2	Sequence 2, Appl
45	36	37.5	524	1	US-08-484-355-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-250-740-33
Sequence 33, Application US/08250740
Patent No. 5685240
GENERAL INFORMATION:
APPLICANT: Schuchman, Edward H.
APPLICANT: Desnick, Robert J.
TITLE OF INVENTION: Acid Sphingomyelinase Gene and Diagnosis
TITLE OF INVENTION: of Niemann-Pick Disease
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/250,740
FILING DATE: 27-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30742
REFERENCE/DOCKET NUMBER: 6923-038
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 629 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-250-740-33

Query Match 45.8%; Score 44; DB 1; Length 629;
Best Local Similarity 71.4%; Pred. No. 35;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 3 VLVLALAVALLAVGDS 16
||| ||||| ||||| ||
DB 36 VLALALALALSD 49

RESULT 2
US-07-695-472B-2
; Sequence 2, Application US/07695472B
; Patent No. 5773278
; GENERAL INFORMATION:
; APPLICANT: Schuchman, Edward H.
; APPLICANT: Desnick, Robert J.
; TITLE OF INVENTION: The Acid Sphingomyelinase Gene and
; TITLE OF INVENTION: Diagnosis of Niemann-Pick Disease
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07695,472B
; FILING DATE: 19910503
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6923-014
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 7908664/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-07-695-472B-2

Query Match 45.8%; Score 44; DB 1; Length 629;
Best Local Similarity 71.4%; Pred. No. 35;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VLVLALAVAGD 16
||| | | | | | | | |
Db 36 VLALALALALSD 49

RESULT 3
US-08-446-137B-2
; Sequence 2, Application US/08446137B
; Patent No. 6162903
; GENERAL INFORMATION:
; APPLICANT: Trowern, Angus R.
; APPLICANT: Atkinson, Anthony
; APPLICANT: Murphy, Jonathan P.
; APPLICANT: Laurence, Oliver S.
; APPLICANT: Duggleby, Clive J.
; TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED
; TITLE OF INVENTION: FROM L PROTEIN AND THEIR USES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA

ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08446,137B
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcmasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.406
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1027 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-446-137B-2

Query Match 45.8%; Score 44; DB 4; Length 1027;
Best Local Similarity 50.0%; Pred. No. 61;
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 2 RVLVLALAVAGDGSN 19
::: ||| | : || | : |
Db 41 KLLMALAGAIIVGGAN 58

RESULT 4
US-08-391-743A-2
; Sequence 2, Application US/08391743A
; Patent No. 5843705
; GENERAL INFORMATION:
; APPLICANT: DiIullo, Paul A.; Meade, Harry; Cole, Edward S.
; TITLE OF INVENTION: TRANSGENETICALLY PRODUCED ANTITHROMBIN III
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08391,743A
; FILING DATE: 21-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Paul Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: TCI-045
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid

; APPLICANT: Natori, Shunji

;; TITLE OF INVENTION: NEW PROTEASE
;; FILE REFERENCE: 32290-144749
;; CURRENT APPLICATION NUMBER: US/09/120.365
;; EARLIER FILING DATE: 1998-07-22
;; EARLIER APPLICATION NUMBER: JP 9-333 474
;; EARLIER FILING DATE: 1997-11-18
;; NUMBER OF SEQ ID NOS: 101
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 75
;; LENGTH: 361
;; TYPE: PRT
;; ORGANISM: Aleurain
US-09-120-365-75

Query Match 41.7%; Score 40; DB 3; Length 361;
Best Local Similarity 54.5%; Pred. No. 77;
Matches 12; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

QY 2 RVLVLAV----ALAVDGSN 19
|||:||||| 1:| 1:
DB 5 RVLALAVLAATAVAVASSS 26

RESULT 8
US-08-676-166A-2
; Sequence 2, Application US/08676166A
; Patent No. 5955270
; GENERAL INFORMATION:
; APPLICANT: Radford, Alan
; APPLICANT: Parish, John H.
; TITLE OF INVENTION: EXPLOITATION OF THE CELLULASE COMPLEX OF
; TITLE OF INVENTION: NEUROSPORA
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,166A
; FILING DATE: 15-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1321-1-002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 516 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-676-166A-2

Query Match 41.7%; Score 40; DB 2; Length 516;
Best Local Similarity 42.9%; Pred. No. 1.2e+02;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MRVLVALAVAVDGSNLG 21
|| :|| :|| :|| :|| :||

DB 1 MRASLAFSLAAAVAGGQAG 21

RESULT 9
US-08-477-396A-4
; Sequence 4, Application US/08477396A
; Patent No. 5872235
; GENERAL INFORMATION:
; APPLICANT: Chen, Lan Bo
; APPLICANT: Bao, Shideng
; APPLICANT: Liu, Yuan
; TITLE OF INVENTION: A NOVEL TUMOR MARKER AND NOVEL METHOD OF
; TITLE OF INVENTION: ISOLATING SAME
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,396A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/146,488
; FILING DATE: 29-OCT-1993
; APPLICATION NUMBER: US 08/448,388
; FILING DATE: 28-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/12502
; FILING DATE: 31-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Helene, Holliday C.
; REGISTRATION NUMBER: 34,346
; REFERENCE/DOCKET NUMBER: DFCI-333BX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-2290
; TELEFAX: (617) 451-0313
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 777 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-477-396A-4

Query Match 41.7%; Score 40; DB 2; Length 777;
Best Local Similarity 56.2%; Pred. No. 1.8e+02;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 VLALAVAVDGSNL 20
:||||:||||:|
DB 7 ILALALALGPAVTL 22

RESULT 10
US-09-040-738-2
; Sequence 2, Application US/09040738
; Patent No. 6207374
; GENERAL INFORMATION:
; APPLICANT: Sampson et al.
; TITLE OF INVENTION: Tuberos Sclerosis 2 Gene and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Kathleen M. Williams, Banner & Witcoff,
ADDRESSEE: Ltd.
STREET: One Financial Center
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 6.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040, 738
FILING DATE: Concurrently herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB9326470.3
FILING DATE: 24-December-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB9411900.5
FILING DATE: 14-June-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB94/02823
FILING DATE: 23-December-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/652,426
FILING DATE: 30-May-1996
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Kathleen M.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: 3265/73963
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-345-9100
TELEFAX: 617-345-9111
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1784 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-040-738-2

Query Match 41.7% Score 40; DB 4; Length 1784;
Best Local Similarity 50.0%; Pred. No. 4.7e+02;
Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 2 RVLVLALAVLAVGDSNLG 21
DB 98 RHAVLALLKAIVQGGERLG 117

RESULT 11
US-08-738-172-4
Sequence 4, Application US/08738172
Patent No. 5939257
GENERAL INFORMATION:
APPLICANT: Szasz, Joseph
APPLICANT: Davis, Maria
TITLE OF INVENTION: THERMOSTABLE ALKALINE PHOSPHATASES
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage

COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/738,172
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/005,965
FILING DATE: October 27, 1995
APPLICATION NUMBER: 08/465,003
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/575,354
FILING DATE: December 20, 1995
APPLICATION NUMBER: 08/240,158
FILING DATE: May 10, 1994
APPLICATION NUMBER: 08/229,329
FILING DATE: April 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 222/158
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-738-172-4

Query Match 40.6% Score 39; DB 2; Length 455;
Best Local Similarity 44.4%; Pred. No. 1.4e+02;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 MRVLVALAVLAVGDS 18
DB 1 MRWILLFLVLLGVGEA 18

RESULT 12
US-08-484-493-12
Sequence 12, Application US/08484493
Patent No. 5728381
GENERAL INFORMATION:
APPLICANT: Wilson, Peter J
APPLICANT: Morris, Charles P
APPLICANT: Anson, Donald S
APPLICANT: Occhiodoro, Teresa
APPLICANT: Bielicki, Julie
APPLICANT: Clements, Peter R
APPLICANT: Hopwood, John J
TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
TITLE OF INVENTION: IDURONATE 2-SULFATASE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/484,493
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION NUMBER: 991,973
FILING DATE: 17-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 84162
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 507 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-493-12

Query Match 40.6%; Score 39; DB 1; Length 507;
Best Local Similarity 52.6%; Pred. No. 1.6e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 2 RVLVLALAVAGDGSNL 20
| | | | | | | | | | | | | | | | | | | | | |
Db 5 RSLLLAAGLAVARPPNI 23

RESULT 13
US-08-484-494-12
Sequence 12, Application US/08484494
Patent No. 5798239
GENERAL INFORMATION:
APPLICANT: Wilson, Peter J
APPLICANT: Morris, Charles P
APPLICANT: Anson, Donald S
APPLICANT: Occhiodoro, Teresa
APPLICANT: Bielicki, Julie
APPLICANT: Clements, Peter R
APPLICANT: Hopwood, John J
TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,494
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 991,973
FILING DATE: 17-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 84162
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343

TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 507 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-494-12

Query Match 40.6%; Score 39; DB 1; Length 507;
Best Local Similarity 52.6%; Pred. No. 1.6e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 2 RVLVLALAVAGDGSNL 20
| | | | | | | | | | | | | | | | | | | | | |
Db 5 RSLLLAAGLAVARPPNI 23

RESULT 14
US-08-345-212-12
Sequence 12, Application US/08345212
Patent No. 5932211
GENERAL INFORMATION:
APPLICANT: Wilson, Peter J
APPLICANT: Morris, Charles P
APPLICANT: Anson, Donald S
APPLICANT: Occhiodoro, Teresa
APPLICANT: Bielicki, Julie
APPLICANT: Clements, Peter R
APPLICANT: Hopwood, John J
TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/345,212
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 991,973
FILING DATE: 17-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 84162
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 507 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-345-212-12

Query Match 40.6%; Score 39; DB 2; Length 507;
Best Local Similarity 52.6%; Pred. No. 1.6e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 RVLVLAVALAVGDSNL 20
| | | | | | | | | |
Db 5 RSLLLAAGLAVARPPNI 23

RESULT 15
US-09-249-003-12
; Sequence 12, Application US/09249003
; Patent No. 6153188
; GENERAL INFORMATION:
; APPLICANT: Wilson, Peter J
; APPLICANT: Morris, Charles P
; APPLICANT: Anson, Donald S
; APPLICANT: Occhiodoro, Teresa
; APPLICANT: Bielicki, Julie
; APPLICANT: Clements, Peter R
; APPLICANT: Hopwood, John J
; TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
; IDURONATE 2-SULFATASE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/09/249,003
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/991,973
; FILING DATE: 17-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8416Z
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 507 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-249-003-12

Query Match 40.6%; Score 39; DB 4; Length 507;
Best Local Similarity 52.6%; Pred. No. 1.6e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 RVLVLAVALAVGDSNL 20
| | | | | | | | | |
Db 5 RSLLLAAGLAVARPPNI 23

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2001, 21:23:16 ; Search time 1155.53 Seconds
(without alignments)
604.820 Million cell updates/sec

Title: US-09-426-776-11

Perfect score: 80

Sequence: 1 attcacatccaccagccatg.....gggaccagtcacacttgggg 80

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*
44: gb_est44:*
45: gb_est45:*
46: gb_est46:*
47: gb_est47:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estp11:*
73: em_estp12:*
74: em_estp13:*
75: em_estp14:*
76: em_estp15:*
77: em_estp16:*
78: em_estp17:*
79: em_estp18:*
80: em_estp19:*
81: em_estp110:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est33:*
111: gb_est34:*
112: gb_est35:*
113: gb_est36:*
114: gb_est37:*
115: gb_est38:*
116: gb_est39:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	33.6	42.0	489	144	BF156722	BF156722 fl56a04.y
2	33.6	42.0	502	144	BF157180	BF157180 fl35g04.y
3	33.6	42.0	559	111	AW133798	AW133798 fl12a02.y
4	33.6	42.0	633	144	BF156786	BF156786 fl56a07.y
5	33.6	42.0	643	144	BF156818	BF156818 fl57c10.y
6	33.6	42.0	718	144	BF157814	BF157814 fl37a10.y
7	33.6	42.0	727	144	BF156997	BF156997 fl59e07.y
8	33.6	42.0	727	144	BF158524	BF158524 fl33g04.y
9	33.6	42.0	736	144	BF156434	BF156434 fl52a06.y
10	33.6	42.0	763	111	AW153678	AW153678 fl24e02.y
11	33.6	42.0	763	111	AW175005	AW175005 fl31d10.y
12	33.6	42.0	764	111	AW153397	AW153397 fl20e02.y
13	33.4	41.8	614	111	AW128094	AW128094 fl06f12.y
14	32.6	40.8	740	144	BF157419	BF157419 fl38e06.y
15	32.4	40.5	667	114	AW343221	AW343221 fl74g10.y
16	32.2	40.3	796	226	AZ697043	AZ697043 RPCI-23-2
17	32	40.0	304	144	BF156771	BF156771 fl56f01.y
18	32	40.0	415	114	AW343413	AW343413 fl77c08.y


```

/db_xref="taxon:7955"
/clone_lib="Sugano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/notes="Vector: pME18S-FL3; Site_1: DraIII (CACTGTCTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCTTCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTCTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science) and kindly donated by Dr. Koichi Kawakami. Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end primer CGACCTGCAGCTCGACACA."
BASE COUNT      125 a 129 c 119 g 129 t
ORIGIN

Query Match      42.0%; Score 33.6; DB 144; Length 502;
Best Local Similarity 66.7%; Pred. No. 0.55;
Matches 48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 7 atccaccagcattggtgttctactagcttctgtgtgctctcagtggtgggacc 66
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 AACCCAGCCATGAGGCTGTGTGCTGCTTGCCTTGCCTGCTAGTCCCTCGCATCTCAAC 62
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 67 agtccaaacttg 78
   | | | | | | | |
Db 63 AATTCACCTTG 74

RESULT 3
LOCUS AW133798 559 bp mRNA EST 27-OCT-1999
DEFINITION f112a02.v1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone 260906 5', similar to FR:093605 093605 VITELLOGENIN ;, mRNA sequence.
ACCESSION AW133798.1 GI:6135405
VERSION AW133798
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE 1 (bases 1 to 559)
AUTHORS Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S., Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T., Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B., Ritter,E., Bowers,Y., Wylie,T., Waterston,R. and Willson,R.
TITLE WashU Zebrafish EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Contact: S. L. Johnson
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA Sequencing by: Washington University Genome Sequencing Center
Seq primer: T3 ET from Amersham
High quality sequence stop: 527.
Location/Qualifiers
1..559
/organism="Danio rerio"
/strain="AB"
FEATURES
source

```

```

/db_xref="taxon:7955"
/clone_lib="Sugano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/notes="Vector: pME18S-FL3; Site_1: DraIII (CACTGTCTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCTTCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTCTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science) and kindly donated by Dr. Koichi Kawakami. Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end primer CGACCTGCAGCTCGACACA."
BASE COUNT      144 a 151 c 128 g 136 t
ORIGIN

Query Match      42.0%; Score 33.6; DB 111; Length 559;
Best Local Similarity 66.7%; Pred. No. 0.56;
Matches 48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 7 atccaccagcattggtgttctactagcttctgtgtgctctcagtggtgggacc 66
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4 AACCCAGCCATGAGGCTGTGTGCTGCTTGCCTTGCCTGCTAGTCCCTCGCATCTCAAC 63
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 67 agtccaaacttg 78
   | | | | | | | |
Db 64 AGATGACCTTG 75

RESULT 4
LOCUS BF156786 633 bp mRNA EST 30-OCT-2000
DEFINITION f156g07.v1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone 3817932 5', similar to TR:093605 093605 VITELLOGENIN ;, mRNA sequence.
ACCESSION BF156786.1 GI:11051974
VERSION BF156786
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE 1 (bases 1 to 633)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Willson,R.
TITLE WashU Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrfish@watson.wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA Sequencing by: Washington University Genome Sequencing Center
distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: T3 ET from Amersham
High quality sequence stop: 496.
Location/Qualifiers

```

source

```

1. .633
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone_lib="Sugano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/notes="Vector: pME18S-FL3; Site 1: DraIII (CACTGTGTG);
Site 2: DraIII (CACCATGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCCCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[GTGTGGCCCTACTGG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end
primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      189 a 159 c 141 g 144 t
ORIGIN

```

```

Query Match      42.0%; Score 33.6; DB 144; Length 633;
Best Local Similarity 66.7%; Pred. No. 0.57;
Matches 48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 7 atccaccacgatgaggtgttactagctcttctgtgctctgcagtgagggaacc 66
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34 AACCCACGCCATGAGAGCTGTGTGCTTGCCTGACTGTAGCCCTGCGGAGTCAAC 93

QY 67 agtccaaacttg 78
      || |||||
Db 94 AGATGACCTTG 105

```

RESULT 5

```

BF156818
LOCUS      BF156818      643 bp      mRNA      EST      30-OCT-2000
DEFINITION fl157c10.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            381747 5', similar to TR:093605 093605 VITELLOGENIN ;, mRNA
            sequence.
ACCESSION  BF156818
VERSION    BF156818.1 GI:11052006
KEYWORDS   EST.
SOURCE     zebrafish.
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE  1 (bases 1 to 643)
AUTHORS    Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
TITLE      WashU zebrafish EST Project 1998
JOURNAL    Unpublished (1998)
COMMENT    Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center Clone
            distribution information can be found through the I.M.A.G.E.
            Consortium/LLNI, send email to: info@image.llnl.gov

```

```

Seq primer: T3 ET from Amersham
High quality sequence stop: 531.

```

FEATURES

```

source
1. .643
Location/Qualifiers
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone_lib="Sugano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/notes="Vector: pME18S-FL3; Site 1: DraIII (CACTGTGTG);
Site 2: DraIII (CACCATGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCCCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[GTGTGGCCCTACTGG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTTCTGCTCTAAAGCTGCG and 3' end
primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      169 a 167 c 154 g 153 t
ORIGIN

```

```

Query Match      42.0%; Score 33.6; DB 144; Length 643;
Best Local Similarity 66.7%; Pred. No. 0.58;
Matches 48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 7 atccaccacgatgaggtgttactagctcttctgtgctctgcagtgagggaacc 66
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3  AACCCACGCCATGAGAGCTGTGTGCTTGCCTGACTGTAGCCCTGCGGAGTCAAC 62

QY 67 agtccaaacttg 78
      || |||||
Db 63 AGATGACCTTG 74

```

RESULT 6

```

BF157814
LOCUS      BF157814      718 bp      mRNA      EST      30-OCT-2000
DEFINITION fl137a10.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
            3815731 5', similar to TR:093605 093605 VITELLOGENIN ;, mRNA
            sequence.
ACCESSION  BF157814
VERSION    BF157814.1 GI:11053013
KEYWORDS   EST.
SOURCE     zebrafish.
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE  1 (bases 1 to 718)
AUTHORS    Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
TITLE      WashU zebrafish EST Project 1998
JOURNAL    Unpublished (1998)
COMMENT    Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA

```

Sequencing by: Washington University Genome Sequencing Center
 Clone distribution information can be found through the I.M.A.G.E.
 Consortium/LLNL, send email to: info@image.llnl.gov
 Seq primer: T3 ET from Amersham
 High quality sequence stop: 485.

FEATURES

1. 718
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone="3815731"
/clone_lib="Sugano Kawakami zebrafish
/sex="mixed (one male and one female,
unfertilized eggs)"
/dev_stage="adult"

/dev_stage="DHI0B (phage resistant)"
 /lab_host="DHI0B (phage resistant)"
 /note="Vector: pME18S-Fl3; Site1: DraIII (CACTGTGTG);
 Site2: DraIII (CACCATGTG); 1st strand cDNA was primed
 with an oligo(dT) primer [ATGCGCCCTTTTTTTTTTTT];
 double-stranded cDNA was ligated to a DraIII adaptor
 [TGTGGCCACTGCG], digested and cloned into distinct DraIII
 sites of the pME18S-Fl3 vector (5' site CACTGTGTG, 3' site
 CACCATGTG). XhoI should be used to isolate the cDNA
 insert. Size selection was performed to exclude fragments
 <1.5kb. Library constructed by Dr. Sumio Sugano
 (University of Tokyo Institute of Medical Science) and
 kindly donated by Dr. Koichi Kawakami. Custom primers for
 sequencing: 5' end primer CTTGTGCTCTAAAGCTGCG and 3' end
 primer CCACCTGCAGCTCGAGCA. "
 183 c 169 a 171 t 1 others

	BASE COUNT	ORIGIN
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0
38	0	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	0
61	0	0
62	0	0
63	0	0
64	0	0
65	0	0
66	0	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0
79	0	0
80	0	0
81	0	0
82	0	0
83	0	0
84	0	0
85	0	0
86	0	0
87	0	0
88	0	0
89	0	0
90	0	0
91	0	0
92	0	0
93	0	0
94	0	0
95	0	0
96	0	0
97	0	0
98	0	0
99	0	0
100	0	0

Query Match 42.0%; Score 33.6; DB 144; Length 718;
Best Local Similarity 56.7%; Pred. NO. 0.59;
Matches 48; Conservative 0; Mismatches 24; Indels 0;

QY	7	atccaccagccatgagggctcttgatctactagctcttctgctgctctcagctgggggacc	66
Db	3	AACCCACCAGCCATGAGAGCTGTTGTGCTTGCCCTGACTCTGCTGGCGAGTCAAC	62
QY	67	agtcccaacttgg	78
Db	63	AGATGAACCTTG	74

RESULT 7
BF156997

LOCUS	BF156997	727 bp	EST	30-OCT-2000
DEFINITION	f159e07.y1 Sugano kawakami zebrafish DRB Danio rerio cDNA clone 3818197 5' similar to TR:093605 O93605 VITELLOGENIN :. mRNA			

sequence.
ACCESSION BF156997
VERSION BF156997.1 GI:11052186
KEYWORDS EST.

SOURCE	zebrafish.
ORGANISM	Danio rerio
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasbora; Danio.

REFERENCE

AUTHORS

Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wallis, J.

1 (pages 1 to 727)

Cyprinodontes; Cyprinidae; Rasbora; Danio.

TITLE Washu Zebrafish EST Project 1998

JOURNAL
COMMENT
Unpublished (1998)
Contact: Stephen L. Johnson

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel.: 314 286 1800

Fax: 314 286 1810
 Email: zbraflshwatson.wustl.edu
 Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
 Sequencing by Washington University Genome Sequencing Center
 Distribution information can be found through the I.M.A.G.E.

FEATURES

```

/organism="Danio rerio"
/strain="AB"
/ob_xref="taxon:7955"
/clone="381819"
/gclone.lib="Sugano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/notes="Vector: pME18S-FU3; Site_1: DraIII (CACTA

```

Site 2: DraIII (CACCATGTG); lnt strand cDNA was primed with an oligo(dT) primer [ATGTCGGCTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [ATGTCGGCTACTGCG], digested and cloned into distinct DraIII sites of the pMEL18-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5 kb. Library constructed by Dr. Sumito Sugano (University of Tokyo Institute of Medical Science) and kindly donated by Dr. Koichi Kawakami. Custom primers for sequencing: 5' end primer CTTTCGTCTCTAAAGCTGCG and 3' end primer CACATCGTACGTCCGACACAC.

BASE COUNT	199 a	185 c	170 g	173 t
ORIGIN				

Query Match 42.0%; Score 33.6; DB 144; Length 727;
Best Local Similarity 66.7%; Pred. No. 0.59;

	Matches	48:	Conservative	0:	Mismatches	24:	Indels	0:	Gaps	0:
Qy	7	atccacacgcatcagggtgctgtgactagctctctgctgctctctgcagtgagggaacc	66							
Db	3	AAACCCAGCAGCATAGAGCTGTTGTGCTTCCTGACTGTAGCCCTCGTGCAGTCAAC	62							
Qy	67	agtcacaaacttgg	78							

RESULT 8

[illegible]

ACCESSION	BF158524	sequence.
VERSION	BF158524.1	GI:11053728
KEYWORDS	EST.	zebrafish.
SOURCE		

ORGANISM

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Actinopterygii: Neopterygii: Teleostei: Euteleostei; Ostariophysi: Cypriniformes: Cyprinidae; Rasbora: Danio.

REFERENCE AUTHORS

S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurr, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

TITLE	WashU Zebrafish EST Project 1998
JOURNAL	Unpublished (1998)
COMMENT	Contact: Stephen I. Johnson

WashU Zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center Clontech
distribution information can be found through the I.M.A.G.E.

distribution information can be found through the I.M.A.G.E.
Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: T3 ET from Amersham
High quality sequence stop: 518.
Location/Qualifiers
1. .736
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone="3817258"
/clone lib="Suzano, Kawakami zebrafish DBA"

```

/clone_lib="Sugano kawakami zebraishn DRA
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/notes=vector: pME18S-FL3; Site.1: DraIII (CACTGTCGTG);
Site.2: DraIII (CACCATGTCG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCCATTTTTTTTTTTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[GTGTGGCCATGTCG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTCG, 3' site
CACCATGTC). xhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and

```

199 a 185 c 178 g 174 t

ch
l Similarity 42.0%; Score 33.6; DB 144; Length 736;
48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

tccaccagcctagggtgcttactagtctcttgctggcttcgcagtaggggaac 66
| | | | | | | | | | | | | | | | | | | | |
TACCACGCGCATGAGAGCTGTGTGCTTGCCCTGACTGTGACCCCTCGTGCAGAGTC AAC 62

gtcccaactgg 78
| | | | |
GATGACCTTG 74.

AW153678 763 bp mRNA 03-NOV-1999
f124e02.y1 Sugano Kawakami zebrafish DRB Danio rerio cDNA clone
2602106 5' similar to TR:093605 O93605.VITELLOGENIN ;, mRNA
sequence.
AW153678
AW153678.1 GI:6201578
EST.
zebrafish.
Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasbora; Danio.

1. (bases 1 to 763)

Sugano, S., Kawakami, K., Johnson, S., Li, F., Marra, M., Eddy, S., Hillier, L., Clifton, S., Allen, M., Gibbons, M., Jost, S., Kucaba, T., Martin, J., Paep, D., Steptoe, M., Underwood, K., Theising, B., Ritter

